1999

Heteroscedastic factor analysis

Sock-Cheng Lewin-Koh
Iowa State University

Heteroscedastic factor analysis

by

Sock-Cheng Lewin-Koh

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Major Professor: Yasuo Amemiya

Iowa State University
Ames, Iowa

1999

UMI Number: 9940221

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Sock-Cheng Lewin-Koh

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1
1.1 The Model 1
1.2 Literature Review 5
CHAPTER 2. DIAGNOSTIC PROCEDURE 7
2.1 An Overview 7
2.2 An Example 9
CHAPTER 3. ESTIMATION PROCEDURES 15
3.1 An Overview 15
3.2 Model-fitting Procedures 16
3.3 Factor Score Estimation 23
CHAPTER 4. ASYMPTOTIC RESULTS 26
4.1 Effect of Heteroscedasticity on the Standard Procedure 26
4.2 Asymptotic Properties of the Heteroscedastic Model Estimators 29
CHAPTER 5. SIMULATION STUDY 36
5.1 Simulation Study I 36
5.2 Simulation Study II 51
5.3 Overall Recommendation 59
APPENDIX. SOME RESULTS 61
BIBLIOGRAPHY 67

CHAPTER 1. INTRODUCTION

1.1 The Model

Factor analysis is a statistical method for modeling and analyzing multivariate data. This method has been used widely in the behavioural sciences for a century. But, it earned recognition as a useful statistical procedure for modeling and inference only in recent years, coinciding with the development of latent variable modeling. The usefulness of factor analysis partly comes from its use of a coherent model, which matches well with many conceptual formulation in applied sciences. The strength of factor analysis as a statistical method lies in the availability of inference procedures that are approximately valid without specifying any distributional form, i.e., that are distribution-free. Factor analysis is concerned with the notion that the structure of multivariate observations can be described by a smaller number of unobservable factors. For a $p \times 1$ vector of observations $\mathbf{Z}_t$ on the the t^{th} individual, the basic factor analysis model is commonly expressed as

$$\mathbf{Z}_t = \boldsymbol{\mu} + \Lambda \mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, ..., n, \tag{1.1}$$

where $\mathbf{f}_t$ is a $k \times 1$ vector of unobservable factors, $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{pt})'$ is a $p \times 1$ vector of unobservable errors, and $\boldsymbol{\mu}$ $(p \times 1)$ and Λ $(p \times k)$ consist of unknown parameters. All interrelationships among the p elements of $\mathbf{Z}_t$ are to be explained by the $k \times 1$ factor $\mathbf{f}_t$. Thus, it is assumed that the p elements of $\boldsymbol{\epsilon}_t$ s are independent, and that the factor $\mathbf{f}_t$ and the error $\boldsymbol{\epsilon}_t$ are independent. It is also standard practice to assume that for each $i = 1, 2, ..., p$, the n individual errors ϵ_{it}, $t = 1, 2, ..., n$, are independent, identically distributed (i.i.d.) with $E(\epsilon_{it}) = 0$ and $Var(\epsilon_{it}) = \psi_{ii,0}$. The factor $\mathbf{f}_t$ can be treated as either random or fixed. If the $\mathbf{f}_t$'s are i.i.d. with $E(\mathbf{f}_t) = \boldsymbol{\mu}_f$ and $Var(\mathbf{f}_t) = \Phi$, then the $\mathbf{Z}_t$'s are i.i.d. with

$$Var(\mathbf{Z}_t) = \Lambda\Phi\Lambda' + \Psi_0, \tag{1.2}$$

where $\Psi_0 = diag(\psi_{11,0}, \psi_{22,0}, ..., \psi_{pp,0})$. In model (1.1), $\mathbf{f}_t$ can be replaced by $\mathbf{C}_0 + \mathbf{C}_1\mathbf{f}_t$ to obtain an equivalent model. To remove this indeterminacy, the so-called errors-in-variables parameterization is used to express the model. This parameterization assumes that, after possible re-ordering of the p variables in $\mathbf{Z}_t$, the model is expressed as

$$\mathbf{Z}_t = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{I}_k \end{pmatrix} \mathbf{f}_t + \boldsymbol{\epsilon}_t. \tag{1.3}$$

This formulation gives an identified model, and also allows straightforward interpretation of $\mathbf{f}_t$ and $\boldsymbol{\beta}_1$. General treatment of factor analysis as a statistical method is given in, for example, Lawley and Maxwell (1963), Anderson (1984), Fuller (1987), and Bollen (1989).

One advantage of the errors-in-variables parameterization (1.3) is that the inference procedures for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ under normality of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$ are valid in large samples for virtually any $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$. See, for example, Anderson and Amemiya (1988), Browne and Shapiro (1988), and Amemiya and Anderson (1990). But, for this result to hold, the model needs to be linear in $\mathbf{f}_t$ and the independence of $\mathbf{f}_t$, ϵ_{1t}, ..., ϵ_{pt} must hold. One recent development in factor analysis and related fields is the use of a model where the relationships between factors and observed variables are nonlinear. See, for example, Kenny and Judd (1984), Ping (1996), Joreskög and Yang (1996), Joreskög and Yang (1997), and Wall and Amemiya (1998). In this dissertation, we consider model (1.3) with possible dependency among $\mathbf{f}_t$, ϵ_{1t}, ..., ϵ_{pt}. In particular, the distribution of the error $\boldsymbol{\epsilon}_t$ for a particular individual t is assumed to depend on some individual characteristics as given by $\mathbf{f}_t$. We propose procedures for exploring such a structure, fitting a model, and making inferences. It is shown that the standard inference procedures for $\boldsymbol{\beta}_1$ in (1.3) are still approximately valid and useful under the dependency between $\mathbf{f}_t$ and ϵ_{it}. But, for estimating and making inferences on the factor values, modeling of the dependence is shown to produce more accurate and useful results.

One way to characterize the dependency of errors on individual characteristics is to consider heteroscedastic error variances. Many areas of statistics deal with a situation where the magnitude of error variability varies among individuals, that is, where the error is dependent on individual characteristics. In other statistical modeling procedures such as linear regression, issues of heteroscedasticity would be addressed by directly modeling it in terms of observed individual characteristics. For multivariate data as in factor analysis, this approach is difficult because there are a large number of error terms, and more importantly, the factors representing individual characteristics are unobservable.

Model (1.1) or (1.3) with the assumption of independence between $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$ and i.i.d. $\boldsymbol{\epsilon}_t$ is termed the *homoscedastic* factor analysis model. The heteroscedasticity of error is often represented as the dependency of the error variances on individual characteristics. Thus, a natural way to express the heteroscedastic structure in factor analysis is to assume that the variances of ϵ_{it} are functions of the $k \times 1$ factor $\mathbf{f}_t$. Then the model becomes

$$
\begin{aligned}
\mathbf{Z}_t &= \boldsymbol{\mu} + \Lambda \mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, ..., n, \\
\psi_{ii,t} &= Var(\epsilon_{it}|\mathbf{f}_t) = g_i^2(\mathbf{f}_t; \boldsymbol{\alpha}), \quad i = 1, 2, ..., p,
\end{aligned}
\tag{1.4}
$$

for some non-negative, scalar-valued function $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$ of $\mathbf{f}_t$ indexed by an unknown parameter vector $\boldsymbol{\alpha}$. The standard deviation of ϵ_{it} is given by $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$. For example, with $k = 1$, we might consider

$$
\begin{aligned}
g_i(f_t; \alpha_{0i}, \alpha_{1i}) &= \alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2, \\
\text{or} \quad g_i(f_t; \mu_i, \lambda_i, c_i) &= c_i(\mu_i + \lambda_i f_t) = c_i E(Z_{it}|f_t),
\end{aligned}
$$

where $\alpha_{0i}, \alpha_{1i}, \alpha_{2i}$ and c_i are additional parameters not related to $\boldsymbol{\mu}$ and Λ in (1.4) and μ_i and λ_i are the i^{th} elements of $\boldsymbol{\mu}$ and Λ. These are examples of polynomial standard deviation functions. Polynomials can describe a broad class of dependency and serve as approximate descriptions in practice. In addition, the polynomial permits a relatively simple model-fitting approach and leads to straightforward testing of heteroscedasticity by testing certain $\boldsymbol{\alpha}$ parameters. Throughout the rest of the thesis, we assume $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$ to be either a polynomial in $\mathbf{f}_t$ or a square root of a polynomial in $\mathbf{f}_t$. We also assume that ϵ_{it}, $i = 1, 2, ..., p$, $t = 1, 2, ..., n$, are conditionally independent given $\mathbf{f}_t$, $t = 1, 2, ..., n$, and that $E(\boldsymbol{\epsilon}_t|\mathbf{f}_t) = 0$. We call model (1.4) the *heteroscedastic* factor analysis model.

Suppose that the $\mathbf{f}_t$'s are i.i.d. with covariance matrix Φ. Then, $Cov(\mathbf{f}_t, \boldsymbol{\epsilon}_t) = E(\mathbf{f}_t E(\boldsymbol{\epsilon}_t|\mathbf{f}_t)) = \mathbf{0}$, and writing

$$\Psi_t = Var(\boldsymbol{\epsilon}_t|\mathbf{f}_t) = diag(\psi_{11,t}, \psi_{22,t}, \cdots, \psi_{pp,t})',$$

we have

$$\begin{aligned} Var(\mathbf{Z}_t) &= \Lambda\Phi\Lambda' + E(\Psi_t) \\ &= \Lambda\Phi\Lambda' + \Psi_0, \end{aligned} \tag{1.5}$$

where

$$\begin{aligned} \Psi_0 &= E(\Psi_t) = diag(\psi_{11,0}, \psi_{22,0}, \cdots, \psi_{pp,0})', \\ \psi_{ii,0} &= E(g_i^2(\mathbf{f}_t; \boldsymbol{\alpha})). \end{aligned}$$

Thus, under the heterocedastic model (1.4), $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$ are uncorrelated, although they are dependent. Also, ϵ_{it}, $i = 1, 2, \ldots, p$, are marginally uncorrelated but dependent. In addition, $\mathbf{Z}_t$, $t = 1, 2, ..., n$, are i.i.d., and $Var(\mathbf{Z}_t)$ in (1.5) has the same form as $Var(\mathbf{Z}_t)$ in (1.2) derived under the homoscedastic model. Hence, the conditional heteroscedastic structure in (1.4) does not affect the form of the marginal second moment of the observations, although it clearly represents the dependency of the conditional error variances on the factor. But, identification and estimation of the $\boldsymbol{\alpha}$ parameters in (1.4) cannot be carried out based only on the first two sample moments of $\mathbf{Z}_t$. This means that we need to develop procedures using information not contained in the first two sample moments. To develop methods that are useful for exploratory modeling and that are valid for a broad class of situations, we do not assume any particular distributional form for $\mathbf{f}_t$ or ϵ_t. In particular, $\mathbf{f}_t$ may be treated as fixed or random with possible dependency over t where (some) observations are taken over time or space. This allows application of our methods for various problems without worrying about the distributional assumptions. The explicit modeling of the heteroscedastic error variances is particularly useful in the estimation or prediction of individual factor $\mathbf{f}_t$. In the next chapter, we will present an informal diagnostic tool that can be used to detect heteroscedasticity and garner some preliminary information about the nature of the heteroscedastic error structure. In Chapter 3, we develop methods for fitting the heteroscedastic error model and for estimating the true factor value. The asymptotic properties of the estimators are derived in Chapter 4. We will show that inference procedures developed in Chapter 3 are asymptotically valid for almost any type of underlying factors, including fixed, correlated, or heteroscedastic factors. We will also examine the effects of error heteroscedasticity on the standard procedure which ignores the heteroscedasticity. Results from simulation studies are presented in Chapter 5 and the appendix gives some formulas useful for computation.

1.2 Literature Review

The focus of pioneering work in factor analysis modeling was on the basic linear factor model (1.1) with normal, homoscedastic errors which are independent of normal, i.i.d. factors. In recent years, some work has been done to extend the basic factor analysis model. Anderson and Amemiya (1988), Browne and Shapiro (1988), and Amemiya and

Anderson (1990) established the robustness of the normal theory asymptotic results to departure from normality. Factor analysis with some form of nonhomogeneous variance has been discussed in only two papers; Meijer and Mooijaart (1996) and Demos and Sentana (1998). Meijer and Mooijaart (1996) considered a special case of model (1.4) and suggested the use of generalized least squares based on the first three sample moments. They gave the asymptotic distribution of their estimators. However, in their simulation study, their estimator showed severe bias and their asymptotic confidence intervals had very poor coverage probabilities (between 42% to 74% for the nominal level 95%).

Demos and Sentana (1998) considered factor analysis model with heteroscedastic factors and homoscedastic errors, and applied the EM algorithm. This model is actually a special case of the basic factor analysis model which allows for any type of factors, as long as the factors are independent of the errors. For their model, the results of Anderson and Amemiya (1988), Browne and Shapiro (1988), and Amemiya and Anderson (1990) apply.

Hasabelnaby (1987) and Sanger and Fuller (1991) discussed estimation of an errors-in-variables model with heteroscedastic errors, which differs considerably from the factor analysis model. Also, they assume that all individual error variances are either known or estimated outside the dataset.

CHAPTER 2. DIAGNOSTIC PROCEDURE

2.1 An Overview

As discussed in Chapter 1, the first two moments of the heteroscedastic model (1.4) have the same form as those of the homoscedastic model (1.1). Hence, the standard factor analysis model fitting procedure using the first two sample moments will not be able to distinguish the two models. As a result, the standard goodness-of-fit testing procedure based on the sample covariance matrix would be expected to have nearly no power to detect the violation of the homoscedasticity. Just as in regression analysis where OLS estimates of the regression parameters remain unbiased even in the presence of heteroscedastic errors, the estimators of $\boldsymbol{\mu}$ and Λ in model (1.1) assuming the homoscedasticity will also be reasonable even if error heteroscedasticity is present. Let the parameterization (1.3) be used. Then the homoscedastic estimators $\tilde{\boldsymbol{\beta}}_0$, $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\Psi}_0$ can be used in our diagnostic method. Since heteroscedasticity in (1.4) concerns $\boldsymbol{\epsilon}_t$ and the relationship of interest is that between the unobservable $\mathbf{f}_t$ and the variance of $\boldsymbol{\epsilon}_t$, it is natural to consider estimated residuals and estimated factor scores for each individual. Since under (1.3)

$$\begin{pmatrix}\mathbf{I}_{p-k}, & -\boldsymbol{\beta}_1\end{pmatrix}\left(\mathbf{Z}_t - \begin{pmatrix}\boldsymbol{\beta}_0 \\ \mathbf{0}\end{pmatrix}\right) = \begin{pmatrix}\mathbf{I}_{p-k}, & -\boldsymbol{\beta}_1\end{pmatrix}\boldsymbol{\epsilon}_t$$

does not involve $\mathbf{f}_t$, we define our $(p-k) \times 1$ estimated residual as

$$\begin{aligned}\tilde{\boldsymbol{\nu}}_t &= \begin{pmatrix}\mathbf{I}_{p-k}, & -\tilde{\boldsymbol{\beta}}_1\end{pmatrix}\left(\mathbf{Z}_t - \begin{pmatrix}\tilde{\boldsymbol{\beta}}_0 \\ \mathbf{0}\end{pmatrix}\right) \\ &= \begin{pmatrix}\mathbf{I}_{p-k}, & -\tilde{\boldsymbol{\beta}}_1\end{pmatrix}\begin{pmatrix}\mathbf{Z}_t - \bar{\mathbf{Z}}\end{pmatrix}. \end{aligned} \tag{2.1}$$

The standard factor score estimator is

$$\tilde{\mathbf{f}}_{hom,t} = \left(\tilde{\Lambda}'\tilde{\Psi}_0^{-1}\tilde{\Lambda}\right)^{-1}\tilde{\Lambda}'\tilde{\Psi}_0^{-1}(\mathbf{Z}_t - \tilde{\boldsymbol{\mu}}), \tag{2.2}$$

where

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix}\tilde{\boldsymbol{\beta}}_0 \\ \mathbf{0}\end{pmatrix}, \quad \tilde{\Lambda} = \begin{pmatrix}\tilde{\boldsymbol{\beta}}_1 \\ \mathbf{I}_k\end{pmatrix}.$$

An alternative formula for $\tilde{\mathbf{f}}_{hom,t}$ is

$$\tilde{\mathbf{f}}_{hom,t} = \begin{pmatrix}\mathbf{0}, & \mathbf{I}_k\end{pmatrix}\left[\mathbf{Z}_t - \tilde{\Psi}_0\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\Psi}_0\tilde{\mathbf{F}})^{-1}\tilde{\boldsymbol{\nu}}_t\right], \tag{2.3}$$

where $\tilde{\mathbf{F}}' = (\mathbf{I}_{p-k}, \quad -\tilde{\boldsymbol{\beta}}_1)$. The second formula for $\tilde{\mathbf{f}}_{hom,t}$ can be used even when $\tilde{\Psi}_0$ is singular. Note that $\tilde{\mathbf{f}}_{hom,t}$ is reasonable for the heteroscedastic model because of the similarity in the first two moments under model (1.1) and model (1.4). Given $\tilde{\mathbf{f}}_{hom,t}$, $\tilde{\Lambda}$ and $\tilde{\boldsymbol{\mu}}$, we can consider a $p \times 1$ estimated residual

$$\tilde{\boldsymbol{\epsilon}}_t = \mathbf{Z}_t - \tilde{\boldsymbol{\mu}} - \tilde{\Lambda}\tilde{\mathbf{f}}_{hom,t} = \tilde{\Psi}_0\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\Psi}_0\tilde{\mathbf{F}})^{-1}\tilde{\boldsymbol{\nu}}_t \tag{2.4}$$

which is a linear function of $\tilde{\boldsymbol{\nu}}_t$.

If $\tilde{\boldsymbol{\beta}}_0$, $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\Psi}_0$ are close to their true values, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and Ψ_0, then

$$\begin{aligned}\tilde{\boldsymbol{\nu}}_t &\approx \begin{pmatrix}\mathbf{I}_{p-k}, & -\boldsymbol{\beta}_1\end{pmatrix}\boldsymbol{\epsilon}_t \\ &= \boldsymbol{\nu}_t, \\ \tilde{\mathbf{f}}_{hom,t} &\approx \mathbf{f}_t + \left(\Lambda'\Psi_0^{-1}\Lambda\right)^{-1}\Lambda'\Psi_0^{-1}\boldsymbol{\epsilon}_t \\ &= \mathbf{f}_t + \begin{pmatrix}\mathbf{0}, & \mathbf{I}_k\end{pmatrix}\left[\boldsymbol{\epsilon}_t - \Psi_0\mathbf{F}\left(\mathbf{F}'\Psi_0\mathbf{F}\right)^{-1}\boldsymbol{\nu}_t\right], \\ \tilde{\boldsymbol{\epsilon}}_t &\approx \Psi_0\mathbf{F}\left(\mathbf{F}'\Psi_0\mathbf{F}\right)^{-1}\boldsymbol{\nu}_t, \end{aligned}$$

where $\mathbf{F}' = (\mathbf{I}_{p-k}, \quad -\boldsymbol{\beta}_1)$. Since $\boldsymbol{\nu}_t$ does not involve $\mathbf{f}_t$, $\tilde{\boldsymbol{\nu}}_t$ and $\tilde{\boldsymbol{\epsilon}}_t$ are approximately uncorrelated with $\tilde{\mathbf{f}}_{hom,t}$ under either the homoscedastic or heteroscedastic model. If error heteroscedasticity is present, we expect it to be reflected in the spread of the residuals when plotted against $\tilde{\mathbf{f}}_{hom,t}$. Hence, the plot of the elements of $\tilde{\boldsymbol{\nu}}_t$ or $\tilde{\boldsymbol{\epsilon}}_t$ against the elements of $\tilde{\mathbf{f}}_{hom,t}$ suggests possible relationship between $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_t$. Also, the squared elements of $\tilde{\boldsymbol{\nu}}_t$ or $\tilde{\boldsymbol{\epsilon}}_t$ can be plotted against the elements of $\tilde{\mathbf{f}}_{hom,t}$ for possible suggestion of the form of $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$. However, we note that $\tilde{\boldsymbol{\nu}}_t$ and $\tilde{\boldsymbol{\epsilon}}_t$ contain linear functions of $\boldsymbol{\epsilon}_t$, and estimating each ϵ_{it} is not possible. If all elements of $\tilde{\boldsymbol{\nu}}_t$ and $\tilde{\boldsymbol{\epsilon}}_t$ exhibit rather homogenous spread across $\tilde{\mathbf{f}}_{hom,t}$, then the standard homoscedastic model should be considered acceptable. On the other hand, if any systematic pattern is found in the plots for any element of $\tilde{\boldsymbol{\nu}}_t$ or $\tilde{\boldsymbol{\epsilon}}_t$, then a model with every error being heteroscedastic should be fitted. In such a case, each error element should be formally examined and tested for heteroscedasticity using the procedures developed in the next chapter. The use of the plots for diagnostic purposes is illustrated using an example in the next section.

2.2 An Example

To illustrate the diagnostic procedure, consider model (1.3) with $k = 1$ and $p = 4$,

$$
\begin{aligned}
\mathbf{Z}_t &= \begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{I}_k \end{pmatrix} f_t + \boldsymbol{\epsilon}_t \\
&= \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ 1 \end{pmatrix} f_t + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}.
\end{aligned} \tag{2.5}
$$

We generated a sample of size 300 with $f_t \sim$ Uniform$(0,3)$ and $\epsilon_{it} \sim$ Normal$(0, \alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2)$ given f_t. The parameter values were

$$\begin{aligned}
\boldsymbol{\beta}_0 &= (1,\ 2,\ 3)', \\
\boldsymbol{\beta}_1 &= (6,\ 5,\ 4)', \\
\boldsymbol{\alpha}_0 &= (\alpha_{01},\ \alpha_{02},\ \alpha_{03},\ \alpha_{04})' = (3,\ 1.5,\ 2,\ 1.5)', \\
\boldsymbol{\alpha}_1 &= (\alpha_{11},\ \alpha_{12},\ \alpha_{13},\ \alpha_{14})' = (1,\ 0,\ 1,\ 0)', \\
\boldsymbol{\alpha}_2 &= (\alpha_{21},\ \alpha_{22},\ \alpha_{23},\ \alpha_{24})' = (2,\ 2,\ 0,\ 0)'.
\end{aligned} \tag{2.6}$$

Thus, ϵ_{4t} is homoscedastic, the variance of ϵ_{3t} is linear in f_t, and those of ϵ_{1t} and ϵ_{2t} are quadratic in f_t. An example of such a situation is calibration with a number of different instruments. The variability of the fourth intrument stays constant for all value of f_t while the accuracy of the first, second, and third instruments decreases with higher values of f_t but at different rates.

Let $(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\beta}}_1, \tilde{\Psi}_0)$ be estimates of $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \Psi_0)$ using a homoscedastic fit. Under model (2.5), the factor and residual estimates in (2.2), (2.1), and (2.4) become

$$\begin{aligned}
\tilde{f}_{hom,t} &= \frac{\sum_{i=1}^{4} \frac{\tilde{\beta}_{1i}(Z_{it} - \tilde{\beta}_{0i})}{\tilde{\psi}_{ii}}}{\sum_{i=1}^{4} \frac{\tilde{\beta}_{1i}^2}{\tilde{\psi}_{ii}}}, \\
\tilde{\nu}_{it} &= Z_{it} - \tilde{\beta}_{0i} - \tilde{\beta}_{1i} Z_{4t}, \quad i = 1, 2, 3, \\
\tilde{\epsilon}_{it} &= Z_{it} - \tilde{\beta}_{0i} - \tilde{\beta}_{1i} \tilde{f}_{hom,t}, \quad i = 1, 2, 3, 4,
\end{aligned} \tag{2.7}$$

with $\tilde{\beta}_{14} = 1$ and $\tilde{\beta}_{04} = 0$. For this sample, the standard large sample χ^2 homoscedastic likelihood ratio goodness of fit test gives a p-value of 0.8273. Thus, the hypothesis that the underlying model is of the form expressed in model (1.1) and (1.2) is not rejected.

Under model (2.5 - 2.6), the error variances of Z_{1t}, Z_{2t} and Z_{3t} vary with f_t, the true value of Z_{4t}. Hence, if it is suspected that error variability depends on some underlying characteristics of the individual as summarized by f_t, then it might be natural to plot

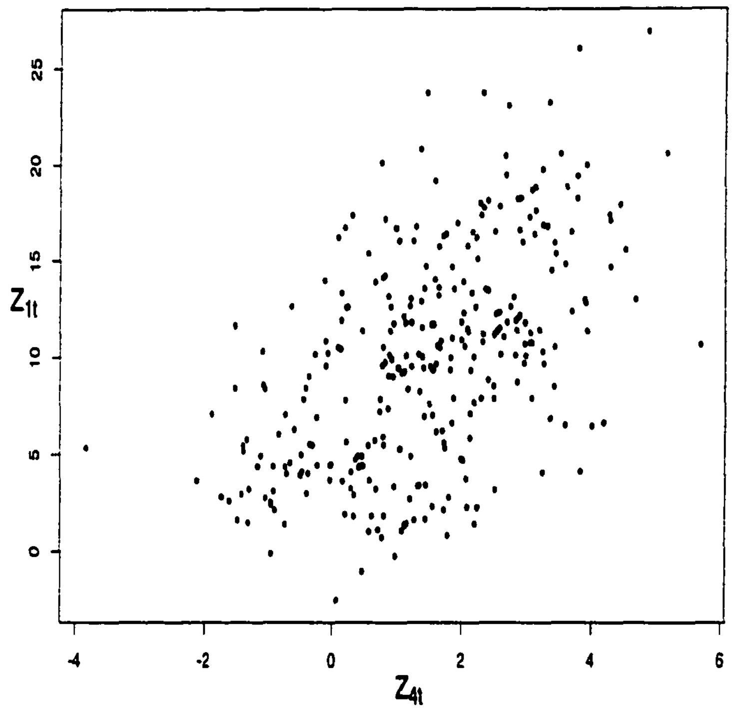

Figure 2.1 Plot of Z_{1t} versus Z_{4t}

Z_{1t}, Z_{2t}, Z_{3t} versus Z_{4t}. As an example, Figure 2.1 is a plot of Z_{1t} versus Z_{4t} for this dataset. There is a slight indication of the heteroscedastic error structure in the plot, but with the presence of the trend, the heteroscedastic pattern is difficult to detect.

As proposed in the previous section, the residuals can be plotted against the estimated factor scores. As an example, Figures 2.2 and 2.3 are plots of $\tilde{\epsilon}_{1t}$ and $\tilde{\epsilon}_{4t}$ versus $\tilde{f}_{hom,t}$, and Figures 2.4 and 2.5 are plots of $\tilde{\epsilon}_{1t}^2$ and $\tilde{\epsilon}_{4t}^2$ versus $\tilde{f}_{hom,t}$. In Figures 2.4 and 2.5, a nonparametric smooth curve using the loess fit is added to assist in perceiving any pattern. The underlying heteroscedastic error structure can be seen in Figures 2.2 and 2.4. Recall that ϵ_{4t} is homoscedastic, which leads to the lack of obvious pattern in Figures 2.3 and 2.5.

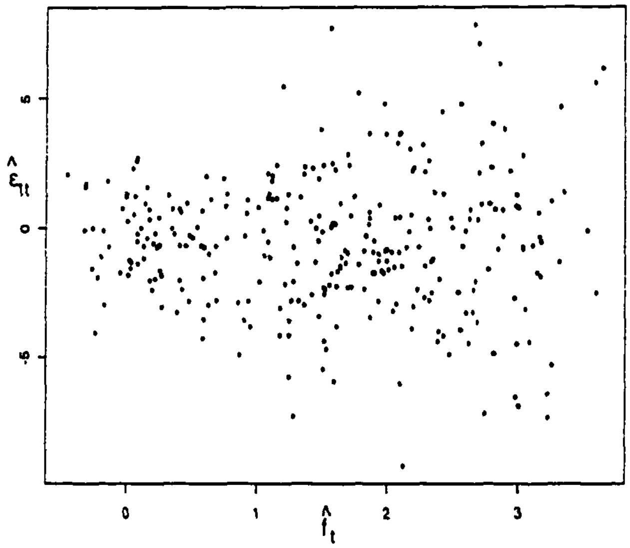

Figure 2.2 Plot of $\tilde{\epsilon}_{1t}$ versus $\tilde{f}_{hom,t}$

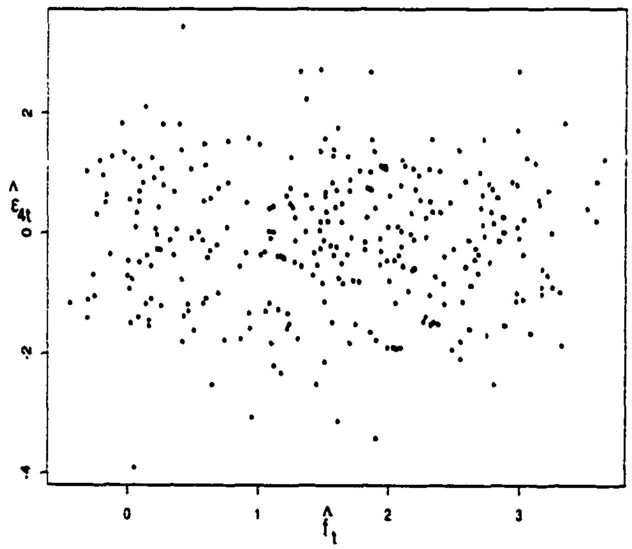

Figure 2.3 Plot of $\tilde{\epsilon}_{4t}$ versus $\tilde{f}_{hom,t}$

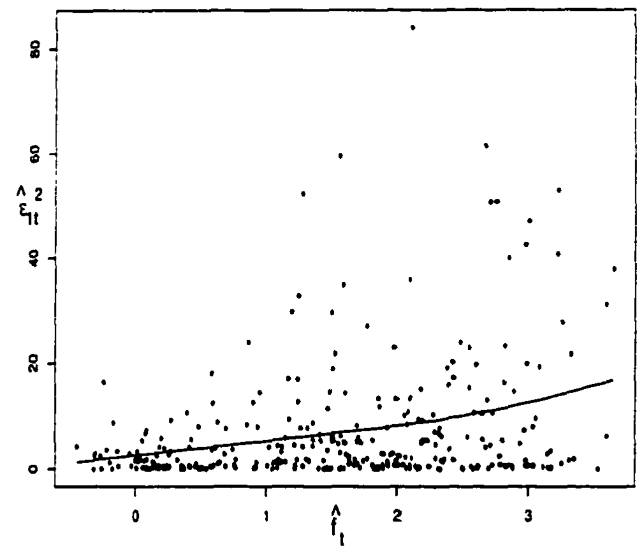

Figure 2.4 Plot of $\tilde{\epsilon}^2_{1t}$ versus $\tilde{f}_{hom,t}$

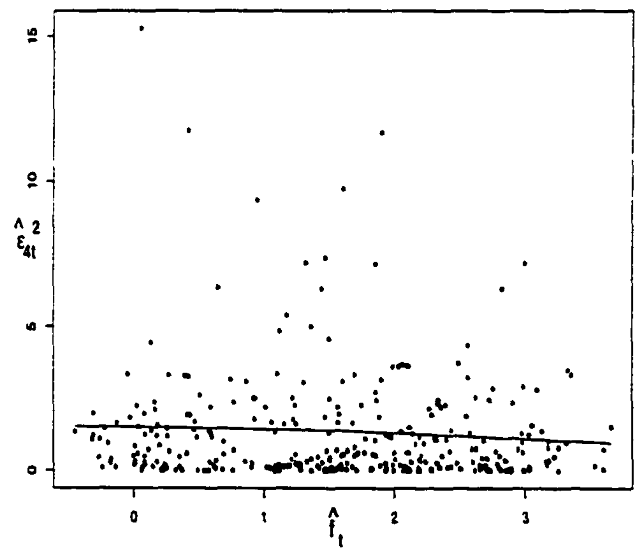

Figure 2.5 Plot of $\tilde{\epsilon}^2_{4t}$ versus $\tilde{f}_{hom,t}$

Besides showing evidence of the relationship between the factor scores and the residuals, the scatterplots also give an indication of the form of the relationship. Thus, for diagnosing heteroscedasticity, the plots involving the two types of residual estimates and $\tilde{f}_{hom,t}$ are recommended over the scatterplots of observed variables Z_{it}'s. As with all graphical procedures, a measure of subjectivity is involved in the assessment of the results. However, the proposed procedure is useful as a tool for preliminary examination of the data. The information culled can then be used in a formal and more rigorous method of model fitting.

CHAPTER 3. ESTIMATION PROCEDURES

3.1 An Overview

Statistical analysis using the heteroscedastic factor analysis model consists of two parts; model-fitting/assessment and factor score estimation. The first part is concerned with estimation and inferences for the error variance parameters of the model and testing for heteroscedasticity. We develop procedures that are valid and useful without specifying the distributional form of the factors and errors. The justification for the procedure based on large-sample theory is given in Chapter 4. The second part, factor score estimation, presents estimators for the value of the underlying factor $\mathbf{f}_t$ for a particular individual t, and gives their approximate standard errors, taking advantage of the fitted heteroscedastic model.

To estimate the heteroscedasticity parameter $\boldsymbol{\alpha}$ in (1.4), we consider a procedure for fitting the whole model, so that an appropriate estimated variance-covariance matrix can be obtained easily. For this, we will be dealing with the moments of the error $\boldsymbol{\epsilon}_t$ of order higher than two. Note that model (1.4) specifies the variance of ϵ_{it} but not the dependency of the other moments of ϵ_{it} on $\mathbf{f}_t$. One way to express the specification of the dependency of $\boldsymbol{\epsilon}_t$ on $\mathbf{f}_t$ in a consistent and coherent fashion is to write model (1.4) as

$$
\begin{aligned}
\mathbf{Z}_t &= \begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{I}_k \end{pmatrix} \mathbf{f}_t + \boldsymbol{\epsilon}_t, \\
\boldsymbol{\epsilon}_t &= (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{pt})', \\
\epsilon_{it} &= g_i(\mathbf{f}_t; \boldsymbol{\alpha})\epsilon_{it}^0, \qquad (3.1)
\end{aligned}
$$

where ϵ_{it}^0 are independent of each other and of $\mathbf{f}_t$, and for each i, ϵ_{it}^0's are i.i.d. with zero mean and unit variance. In the model (3.1), the relationship between $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_t$, as well as their joint distribution are explicitly specified.

We consider model (3.1) and develop estimation and model-fitting procedures without relying on the distributional form of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$. The next section proposes a moment-based approach that is valid for any distributions of i.i.d. $\boldsymbol{\epsilon}_t^0$ and any type of fixed or random $\mathbf{f}_t$ with possible dependency over t.

3.2 Model-fitting Procedures

To identify and fit the heteroscedastic model (3.1), we need to use information not found in the first two sample moments of $\mathbf{Z}_t$. One way to include such information in the analysis is to augment the observation vector $\mathbf{Z}_t$ with some functions of the elements of $\mathbf{Z}_t$. This idea of augmentation was introduced by Kenny and Judd (1984). They were interested in fitting a quadratic or cross-product latent variable model, and suggested using the products of the elements of the observation vector in the estimation procedure. This method was subsequently adopted by several other researchers investigating similar problems, including Jaccard and Wan (1995), Jöreskog and Yang (1996), Ping (1996), Jöreskog and Yang (1997), and Wall and Amemiya (1998). Our problem is not directly related to the polynomial latent variable model discussed in the literature. But, there are similarities between the two, and we can utilize the same idea of augmentation in our model-fitting procedure. Since $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ is polynomic in $\mathbf{f}_t$ or is a square-root of a polynomial in $\mathbf{f}_t$, it is natural to consider pure and mixed powers of $Z_{it} - \bar{Z}_i, i = 1, 2, ..., p$, where Z_{it} is the i^{th} element of $\mathbf{Z}_t$, and $\bar{Z}_i = \frac{1}{n}\sum_{t=1}^{n} Z_{it}$. Let $\mathbf{U}_t$ be the vector of new variables to be added to $\mathbf{Z}_t$ and write the augmented observation vector as

$$\mathbf{Z}_{a,t} = \begin{pmatrix} \mathbf{Z}_t \\ \mathbf{U}_t \end{pmatrix}. \tag{3.2}$$

The first two moments of these additional variables give us the additional information necessary to identify the model. Since the elements of $\mathbf{U}_t$ are pure and mixed powers of Z_{it}, the first two sample moments of $\mathbf{Z}_{a,t}$ include moments of Z_{it} of order three and higher. But, the sample moments of $\mathbf{Z}_{a,t}$ contain some redundant duplicates, because $\mathbf{U}_t$ is a function of $\mathbf{Z}_t$. Our model fitting procedures estimate all the parameters appearing in the (approximate) expectations of the first two sample moments of $\mathbf{Z}_{a,t}$. In obtaining the approximate expectations, we act as if $\bar{Z}_i$ used in $\mathbf{U}_t$ are replaced by $E(Z_{it})$ and $\mathbf{f}_t$'s are i.i.d.. Denote such expectations by

$$\begin{aligned} \boldsymbol{\mu}_a(\boldsymbol{\theta}) &\approx E(\mathbf{Z}_{a,t}), \\ \Xi_{\mathbf{a}}(\boldsymbol{\theta}) &\approx E(\mathbf{Z}_{a,t}\mathbf{Z}'_{a,t}), \\ \boldsymbol{\zeta}(\boldsymbol{\theta}) &= \begin{pmatrix} \boldsymbol{\mu}_a(\boldsymbol{\theta}) \\ vech\, \Xi_{\mathbf{a}}(\boldsymbol{\theta}) \end{pmatrix}, \end{aligned} \tag{3.3}$$

where the matrix operator *vech* lists distinct elements of a symmetric matrix in a column vector, and $\boldsymbol{\theta}$ includes all unknown parameters appearing in $\boldsymbol{\zeta}(\boldsymbol{\theta})$. Since we do not make any assumptions on the distributional form of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$, and since we act as though $\mathbf{f}_t$'s are i.i.d., the moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ need to be estimated as unrestricted parameters and are included in $\boldsymbol{\theta}$ ($E(\epsilon_{it}^0) = 0$ and $Var(\epsilon_{it}^0) = 1$). Let $\boldsymbol{\theta}_{f\epsilon}$ denote the vector of all moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ to be estimated. Then, we can write

$$\boldsymbol{\theta} = \left(\boldsymbol{\beta}'_0,\ \ vec\boldsymbol{\beta}'_1,\ \ \boldsymbol{\alpha}',\ \ \boldsymbol{\theta}'_{f\epsilon}\right)', \tag{3.4}$$

where the matrix operator *vec* stacks the columns of a matrix on top of each other to form a column vector. When we include higher order powers of $\mathbf{Z}_t$ in $\mathbf{U}_t$, the number of higher order moments of $\mathbf{f}_t$ and ϵ_t^0 to be estimated increases and so does the dimension of $\boldsymbol{\theta}_{f_t}$. But, $\mathbf{f}_t$ is $k \times 1$ with $k < p$, and ϵ_{it}^0, $i = 1, 2, ..., p$ are independent, while the number of possible $\mathbf{Z}_t$ moments of a given order is related to p. Thus, the model and parameters are identified (by the counting rule) based on the finite two moments of $\mathbf{Z}_{a,t}$ by choosing an appropriate $\mathbf{U}_t$ so that the number of distinct elements of $\boldsymbol{\zeta}(\boldsymbol{\theta})$ is larger than or equal to the dimension of $\boldsymbol{\theta}$. It is not simple to give a general guideline for all possible cases involving $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ of possibly different order. But, if $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$, $i = 1, 2, ..., p$, are all linear in $\mathbf{f}_t$, inclusion of cross-products $(Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j)$, $i \neq j$, in $\mathbf{U}_t$ suffices for identification. If $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$, $i = 1, 2, ..., p$, are all of order $d_g > 1$, then $\mathbf{U}_t$ containing all moments of $(Z_{it} - \bar{Z}_i)$ up to order d_g is sufficient. In general, mixed power terms, as compared to pure power terms are more helpful for identification and are more stable statistically, because of the independence of ϵ_{it}^0, $i = 1, 2, ..., p$. Also, including much larger number of terms than just needed for identification tends to increase the sampling variability of the parameter estimators in small samples. Throughout, we assume that the corresponding homoscedastic model is identified, i.e., $\frac{(p-k)(p-k+1)}{2} \geq p$, so that k is small compared to p. The homoscedastic model, a special case of the heteroscedastic model, needs to be identified before considering more complex models.

To illustrate the idea of augmentation, consider the following one-factor heteroscedastic model which was also used in the example in Section 2.2,

$$
\begin{aligned}
\mathbf{Z}_t &= \begin{pmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \\ Z_{4t} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ 1 \end{pmatrix} f_t + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}, \qquad (3.5) \\
\epsilon_{it} &= \sqrt{\alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2} \;\; \epsilon_{it}^0.
\end{aligned}
$$

Under this model, two possible additional variables to be included in $\mathbf{U}_t$ for identification and estimation of the $\boldsymbol{\alpha}$-parameters are

$$Y_{it} = (Z_{it} - \bar{Z}_i)^2 \quad \text{and} \quad W_{ij,t} = (Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j).$$

Let $\mathbf{U}_t = (\mathbf{W}_t', \mathbf{Y}_t')'$, where $\mathbf{W}_t'$ and $\mathbf{Y}_t'$ consist of those $W_{ij,t}$ and Y_{it} chosen to be used. Then, the augmented observation vector is $\mathbf{Z}_{a,t} = (\mathbf{Z}_t', \mathbf{U}_t')'$. If only $W_{ij,t}$'s are used, the expectation $\boldsymbol{\zeta}(\boldsymbol{\theta})$ in (3.3) involves polynomials in β_{0i}, β_{1i}, α_{0i}, α_{1i}, α_{2i}, and the first four moments of f_t. But, if Y_{it}'s are used, explicit expressions cannot be found for some of elements of $\boldsymbol{\zeta}(\boldsymbol{\theta})$ (see Appendix). Hence for this model, $W_{ij,t}$ is preferred over Y_{it} for use as additional variables.

In general, let

$$\begin{aligned} \bar{\mathbf{Z}}_a &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{Z}_{a,t}, \\ \mathbf{M}_a &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{Z}_{a,t}\mathbf{Z}_{a,t}'. \end{aligned} \tag{3.6}$$

The moment-based model-fitting procedure is to minimize some distance between $\boldsymbol{\zeta}(\boldsymbol{\theta})$ and $(\bar{\mathbf{Z}}_a', (vech\mathbf{M}_a)')'$, and we consider two types of distance measures. In presenting the methods, we act as though $\mathbf{f}_t$'s are i.i.d.. As shown in Chapter 4, the resulting estimation and inference procedures are valid for other types of $\mathbf{f}_t$.

The first distance measure considered here is a weighted least squares measure. We write

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{Z}_{a,t} \\ vech(\mathbf{Z}_{a,t}\mathbf{Z}_{a,t}') \end{pmatrix}, \tag{3.7}$$

so that

$$\begin{pmatrix} \bar{\mathbf{Z}}_a \\ vech\,\mathbf{M}_a \end{pmatrix} = \bar{\mathbf{A}} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{A}_t. \tag{3.8}$$

This form suggests a distribution-free estimator of the covariance matrix of $\bar{\mathbf{A}}$ given by

$$\widehat{\Pi} = \frac{1}{n(n-1)}\sum_{t=1}^{n}(\mathbf{A}_t - \bar{\mathbf{A}})(\mathbf{A}_t - \bar{\mathbf{A}})'. \tag{3.9}$$

Note that $\widehat{\Pi}$ is singular by the augmentation construction of $\mathbf{Z}_{a,t}$. The weighted least squares estimator $\hat{\boldsymbol{\theta}}_{WLS}$ is the value of $\boldsymbol{\theta}$ that minimizes

$$(\bar{\mathbf{A}} - \boldsymbol{\zeta}(\boldsymbol{\theta}))' \, \widehat{\Pi}^{+} \, (\bar{\mathbf{A}} - \boldsymbol{\zeta}(\boldsymbol{\theta})) , \tag{3.10}$$

where $\widehat{\Pi}^{+}$ is the Moore-Penrose generalized inverse of $\widehat{\Pi}$. Note that the singularity in $\widehat{\Pi}$ is the same as the redundancy in $\boldsymbol{\zeta}(\boldsymbol{\theta})$. Thus, a straightforward estimated covariance matrix of $\hat{\boldsymbol{\theta}}_{WLS}$ is

$$\widehat{\mathbf{V}}_{WLS} = (\widehat{\Delta}^{*\prime}\widehat{\Pi}^{+}\widehat{\Delta}^{*})^{-1} , \tag{3.11}$$

where $\widehat{\Delta}^{*} = \frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{\zeta}(\hat{\boldsymbol{\theta}}_{WLS})$.

We note that $\widehat{\Pi}$ in (3.10) involves the fourth order sample moments of $\mathbf{Z}_{a,t}$, and $\mathbf{Z}_{a,t}$ contains powers of elements of $\mathbf{Z}_t$. Thus, some very high order moments of $\mathbf{Z}_t$ are employed in the weighted least squares estimation procedure. Alternatively, our second distance measure involves the use of only the first two sample moments of $\mathbf{Z}_{a,t}$ in the estimation procedure. This measure is related to the likelihood for normal $\mathbf{Z}_{a,t}$'s. Because the augmented observation vector $\mathbf{Z}_{a,t}$ contains powers of $\mathbf{Z}_t$, $\mathbf{Z}_{a,t}$ cannot be normal even if $\mathbf{Z}_t$ is normal. But, we can still use the normal likelihood as a distance measure involving only the first two moments. For this, it is more natural to use the sample covariance matrix than the uncorrected sum of squares $\mathbf{M}_a$ in (3.6). Let

$$\begin{aligned} \mathbf{S}_a &= \frac{1}{n-1}\sum_{t=1}^{n}(\mathbf{Z}_{a,t} - \bar{\mathbf{Z}}_a)(\mathbf{Z}_{a,t} - \bar{\mathbf{Z}}_a)' \\ &= \frac{n}{n-1}\mathbf{M}_a - \frac{n}{n-1}\bar{\mathbf{Z}}_a\bar{\mathbf{Z}}_a', \\ \Sigma_a(\boldsymbol{\theta}) &= \Xi_a(\boldsymbol{\theta}) - \boldsymbol{\mu}_a(\boldsymbol{\theta})\boldsymbol{\mu}_a(\boldsymbol{\theta})'. \end{aligned} \tag{3.12}$$

Then, the pseudo likelihood estimator $\hat{\boldsymbol{\theta}}_{PL}$ minimizes the pseudo likelihood distance

$$(\bar{\mathbf{Z}}_a - \boldsymbol{\mu}_a(\boldsymbol{\theta}))' \, \Sigma_a^{-1}(\boldsymbol{\theta}) \, (\bar{\mathbf{Z}}_a - \boldsymbol{\mu}_a(\boldsymbol{\theta})) + tr(\mathbf{S}_a\Sigma_a^{-1}(\boldsymbol{\theta})) \; + \; log(|\Sigma_a(\boldsymbol{\theta})|) \, . \tag{3.13}$$

One advantage of $\hat{\boldsymbol{\theta}}_{PL}$ is that the existing software packages can be used for computation. However, the approximate covariance matrix of $\hat{\boldsymbol{\theta}}_{PL}$ given by the packages is not valid, since $\mathbf{Z}_{a,t}$ is not normal. If $\mathbf{f}_t$'s are i.i.d., i.e., if $\mathbf{Z}_{a,t}$'s are i.i.d., then a distribution-free estimator of the variance of $(\bar{\mathbf{Z}}_a', (vech\,\mathbf{S}_a)')'$ can be obtained in a straightforward manner. Let

$$\begin{aligned} \mathbf{a}_t &= \begin{pmatrix} \mathbf{Z}_{a,t} \\ vech\,(\mathbf{Z}_{a,t} - \bar{\mathbf{Z}}_a)(\mathbf{Z}_{a,t} - \bar{\mathbf{Z}}_a)' \end{pmatrix}, \\ \bar{\mathbf{a}} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{a}_t\,, \\ \widehat{\Gamma} &= \frac{1}{n(n-1)}\sum_{t=1}^{n} (\mathbf{a}_t - \bar{\mathbf{a}})(\mathbf{a}_t - \bar{\mathbf{a}})'. \end{aligned} \tag{3.14}$$

Combining this $\widehat{\Gamma}$ and the standard normal likelihood theory, we can develop a distribution free asymptotic covariance matrix estimator for i.i.d. $\mathbf{f}_t$ as

$$\widehat{\mathbf{V}}_{PL} = (\widehat{\Delta}'\widehat{\Omega}^{-1}\widehat{\Delta})^{-1}\widehat{\Delta}'\widehat{\Omega}^{-1}\widehat{\Gamma}\widehat{\Omega}^{-1}\widehat{\Delta}(\widehat{\Delta}'\widehat{\Omega}^{-1}\widehat{\Delta})^{-1}, \tag{3.15}$$

where

$$\begin{aligned} \widehat{\Delta} &= \frac{\partial}{\partial \boldsymbol{\theta}'}\begin{pmatrix} \boldsymbol{\mu}_a(\hat{\boldsymbol{\theta}}_{PL}) \\ vech\,\Sigma_a(\hat{\boldsymbol{\theta}}_{PL}) \end{pmatrix}, \\ \widehat{\Omega} &= \begin{bmatrix} \Sigma_a(\hat{\boldsymbol{\theta}}_{PL}) & \mathbf{0} \\ \mathbf{0} & 2\mathbf{K}^{+}(\Sigma_a(\hat{\boldsymbol{\theta}}_{PL}) \otimes \Sigma_a(\hat{\boldsymbol{\theta}}_{PL}))\mathbf{K}^{+\prime} \end{bmatrix}, \end{aligned} \tag{3.16}$$

$$\mathbf{K}^{+} = (\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}', \tag{3.17}$$

and $\mathbf{K}$ is the known matrix such that $vec\,\Sigma_a(\boldsymbol{\theta}) = \mathbf{K}\,vech\,\Sigma_a(\boldsymbol{\theta})$. The asymptotic theory in Chapter 4 shows that $\widehat{\mathbf{V}}_{PL}$ can be used correctly for non-i.i.d. $\mathbf{f}_t$.

In the limit or for very large n, we expect $\hat{\boldsymbol{\theta}}_{WLS}$ to be more efficient than the $\hat{\boldsymbol{\theta}}_{PL}$. But, $\widehat{\Gamma}$ and $\widehat{\Pi}$ involving higher order sample moments, tend to be very variable. Note that $\widehat{\Gamma}$ is used for estimating the approximate variance of $\hat{\boldsymbol{\theta}}_{PL}$ but not for obtaining $\hat{\boldsymbol{\theta}}_{PL}$. On the other hand, $\widehat{\Pi}$ is used both for estimating $\hat{\boldsymbol{\theta}}_{WLS}$ and in its approximate variance. Thus, in finite samples, $\hat{\boldsymbol{\theta}}_{PL}$ can be more efficient than $\hat{\boldsymbol{\theta}}_{WLS}$.

For polynomial $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$, homoscedasticity or no dependency between ϵ_{it} and $\mathbf{f}_t$ corresponds to the zero condition for all $\boldsymbol{\alpha}$-parameters appearing in $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$ except for the intercept. Thus, a part of $\hat{\boldsymbol{\theta}}_{PL}$ and $\widehat{\mathbf{V}}_{PL}$ or $\hat{\boldsymbol{\theta}}_{WLS}$ and $\widehat{\mathbf{V}}_{WLS}$ corresponding to the relevant $\boldsymbol{\alpha}$ parameters can be used to test for the homoscedasticity for ϵ_{it} for all i or each i.

For instance, in the example model (2.5)-(2.6) with error variance given by $Var(\epsilon_{it}) = \alpha_{0i} + \alpha_{1i}f_t + \alpha_{2i}f_t^2$, the heteroscedastic case reduces to the homoscedastic one when $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{0}$. Let $\hat{\boldsymbol{\theta}}$ denote the parameter estimator from either the PL or WLS fit, and let $\widehat{Var}(\hat{\boldsymbol{\theta}})$ denote the corresponding estimated approximate covariance matrix. If our interest is in testing for overall presence of error heteroscedasticity in the model, i.e., testing

$$H_0 : \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{0},$$

we can compare the test statistic

$$\widehat{T} = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \end{pmatrix}' \widehat{\mathbf{V}}_{\boldsymbol{\alpha}}^{-1} \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \end{pmatrix} \tag{3.18}$$

with the χ^2_8 distribution, where $(\hat{\boldsymbol{\alpha}}'_1, \hat{\boldsymbol{\alpha}}'_2)'$ is the 8×1 estimate from the heteroscedastic fit, and $\widehat{\mathbf{V}}_{\boldsymbol{\alpha}}$ is the 8×8 submatrix of $\widehat{Var}(\hat{\boldsymbol{\theta}})$ corresponding to $(\hat{\boldsymbol{\alpha}}'_1, \hat{\boldsymbol{\alpha}}'_2)'$. If the interest lies in detecting the presence of heteroscedasticity in individual error components, then a test of

$$H_0 : \alpha_{1i} = \alpha_{2i} = 0,$$

can be carried out by comparing

$$\widehat{T}_i = \begin{pmatrix} \hat{\alpha}_{1i} \\ \hat{\alpha}_{2i} \end{pmatrix}' \widehat{\mathbf{V}}_i^{-1} \begin{pmatrix} \hat{\alpha}_{1i} \\ \hat{\alpha}_{2i} \end{pmatrix} \tag{3.19}$$

with the χ_2^2 distribution, where $(\hat{\alpha}_{1i}, \hat{\alpha}_{2i})'$ is the 2×1 estimate from the heteroscedastic fit and $\widehat{\mathbf{V}}_i$ is the 2×2 submatrix of $\widehat{Var}(\hat{\boldsymbol{\theta}})$ corresponding to $(\hat{\alpha}_{1i}, \hat{\alpha}_{2i})'$. To see whether the quadratic term is needed in the variance of ϵ_{it}, we test

$$H_0 : \alpha_{2i} = 0,$$

by comparing the test statistic

$$T_{2i} = \frac{\hat{\alpha}_{2i}}{\sqrt{\widehat{V}(\hat{\alpha}_{2i})}} \tag{3.20}$$

with the standard normal distribution, where $\widehat{V}(\hat{\alpha}_{2i})$ is the diagonal element of $\widehat{Var}(\hat{\boldsymbol{\theta}})$ corresponding to $\hat{\alpha}_{2i}$.

3.3 Factor Score Estimation

Estimation of the true value of the underlying factor is often of interest . For example, a financial institution may desire some measurement of an individual's credit worthiness. If $\mathbf{Z}_t$ consists of p measurements on the t^{th} individual relating to his financial standing and one of the factors is credit worthiness, then an estimate of that factor for a specific individual would be a measurement of the individual's credit worthiness. In such situations, some accuracy of the individual estimate, for example, a confidence interval, would be informative. Under the homoscedastic factor analysis model, the appropriate standard error of the factor score estimator is constant for all individuals. Upon fitting the heteroscedastic model, we can obtain an estimated standard error of the factor score estimate that depends on the individual. This individual-specific accuracy of the factor estimator can be informative and useful in practice.

Let the heteroscedastic model (3.1) hold. The usual homoscedastic factor score estimator is as given in (2.2),

$$\tilde{\mathbf{f}}_{hom,t} = \left(\tilde{\Lambda}'\tilde{\Psi}_0^{-1}\tilde{\Lambda}\right)^{-1} \tilde{\Lambda}'\tilde{\Psi}_0^{-1} (\mathbf{Z}_t - \tilde{\boldsymbol{\mu}}), \tag{3.21}$$

where

$$\tilde{\Lambda} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \mathbf{I}_k \end{pmatrix}, \quad \tilde{\boldsymbol{\mu}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_0 \\ \mathbf{0} \end{pmatrix},$$

and $\tilde{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\Psi}_0$ are obtained from fitting the homoscedastic model. The approximate variance of $\tilde{\mathbf{f}}_{hom,t}$ can be estimated by the standard estimator

$$\begin{aligned} \tilde{V}_{HOM} &= \left(\tilde{\Lambda}'\tilde{\Psi}_0^{-1}\tilde{\Lambda}\right)^{-1}. \qquad (3.22) \\ &= \begin{pmatrix} \mathbf{0}, & \mathbf{I}_k \end{pmatrix} \left[\tilde{\Psi}_0 - \tilde{\Psi}_0\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\Psi}_0\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}'\tilde{\Psi}_0\right] \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_k \end{pmatrix}, \end{aligned}$$

where $\tilde{\mathbf{F}}' = (\mathbf{I}_{p-k}, \quad -\tilde{\boldsymbol{\beta}}_1)$, and the second formula can be used with singular $\tilde{\Psi}_0$. An alternative form of (3.21) which can be used with singular $\tilde{\Psi}_0$ is given in (2.3). Note that $\tilde{V}_{HOM}$ is constant for all t. If the heteroscedasticity model holds, $\tilde{\mathbf{f}}_{hom,t}$ using the average weight $\tilde{\Psi}_0$ is still a reasonable estimator, but it may not be the most efficient. More importantly, $\tilde{V}_{HOM}$ may not provide the best estimate of the variability in a particular $\tilde{\mathbf{f}}_{hom,t}$.

With a heteroscedastic model fit, we can obtain an individual error variance estimator

$$\begin{aligned} \hat{\Psi}_t^{(1)} &= diag(\hat{\psi}_{11,t}^{(1)}, \hat{\psi}_{22,t}^{(1)}, \ldots, \hat{\psi}_{pp,t}^{(1)}), \\ \hat{\psi}_{ii,t}^{(1)} &= g_i^2(\tilde{\mathbf{f}}_{hom,t}, \hat{\boldsymbol{\alpha}}), \qquad (3.23) \end{aligned}$$

where $\hat{\boldsymbol{\alpha}}$ is the PL or WLS estimator from the heteroscedastic fit. Then, an individualized estimate of the variance of $\tilde{\mathbf{f}}_{hom,t}$ is

$$\widehat{V}_{HOM,t} = \left(\tilde{\Lambda}'\tilde{\Psi}_0^{-1}\tilde{\Lambda}\right)^{-1} \tilde{\Lambda}'\tilde{\Psi}_0^{-1}\hat{\Psi}_t^{(1)}\tilde{\Psi}_0^{-1}\tilde{\Lambda} \left(\tilde{\Lambda}'\tilde{\Psi}_0^{-1}\tilde{\Lambda}\right)^{-1}. \qquad (3.24)$$

Given the heteroscedastic fit and $\widehat{\Psi}_t^{(1)}$, we can define the heteroscedastic factor score estimator

$$\hat{\mathbf{f}}_{het,t}^{(1)} = \left(\hat{\Lambda}'\hat{\Psi}_t^{(1)^{-1}}\hat{\Lambda}\right)^{-1} \hat{\Lambda}'\hat{\Psi}_t^{(1)^{-1}} (\mathbf{Z}_t - \hat{\boldsymbol{\mu}}), \qquad (3.25)$$

where

$$\hat{\Lambda} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \mathbf{I}_k \end{pmatrix}, \quad \hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ \mathbf{0} \end{pmatrix},$$

and $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_0$ are the PL or WLS heteroscedastic estimators. Then, we can obtain a new estimator of the individual error variance as

$$\begin{aligned} \hat{\Psi}_t^{(2)} &= diag(\hat{\psi}_{11,t}^{(2)}, \hat{\psi}_{22,t}^{(2)}, \ldots, \hat{\psi}_{pp,t}^{(2)}), \\ \hat{\psi}_{ii,t}^{(2)} &= g_i^2(\hat{\mathbf{f}}_{het,t}^{(1)}, \hat{\boldsymbol{\alpha}}), \end{aligned} \tag{3.26}$$

and a new heteroscedastic factor score estimator is

$$\hat{\mathbf{f}}_{het,t}^{(2)} = \left(\hat{\Lambda}'\hat{\Psi}_t^{(2)^{-1}}\hat{\Lambda}\right)^{-1} \hat{\Lambda}'\hat{\Psi}_t^{(2)^{-1}} (\mathbf{Z}_t - \hat{\boldsymbol{\mu}}), \tag{3.27}$$

with an estimated variance

$$\widehat{Var}(\hat{\mathbf{f}}_{het,t}^{(2)}) = \left(\hat{\Lambda}'\hat{\Psi}_t^{(2)^{-1}}\hat{\Lambda}\right)^{-1}. \tag{3.28}$$

This process can be iterated for a few times till the factor score estimates and the variance estimates stabilize, or we can simply use $\hat{\mathbf{f}}_{het,t}^{(2)}$ and $\widehat{Var}(\hat{\mathbf{f}}_{het,t}^{(2)})$. Alternative forms of (3.24) (3.25), (3.27) and (3.28) can be obtained as in (2.3) and (3.22).

The model-fitting procedures presented in Section 3.2 yield consistent estimators for $\boldsymbol{\mu}$, Λ and $\boldsymbol{\alpha}$. If in addition, ϵ_t^0's in model (3.1) are i.i.d. normal variates, then the conditional distribution of the error in a factor score estimator given $\mathbf{f}_t$ is approximately normal. Thus, an estimated standard error based on (3.24) or (3.28) and the standard normal percentiles can be used to construct an approximate confidence interval.

CHAPTER 4. ASYMPTOTIC RESULTS

In Section 4.1 of this chapter, we discuss the effect of error heteroscedasticity on the asymptotic properties of the standard estimator which assumes homoscedasticity. The limiting distributions of the heteroscedastic model estimators are derived in Section 4.2.

4.1 Effect of Heteroscedasticity on the Standard Procedure

The standard procedure in factor analysis assumes that the observation vector $\mathbf{Z}_t$ is normally distributed and the factors and errors are homoscedastic and independent. Anderson and Amemiya (1988) showed that the limiting distribution of the normal maximum likelihood estimator of the factor loading $\boldsymbol{\beta}_1$ is common for a wide class of factor and error distributions. Thus, the asymptotic inferences for $\boldsymbol{\beta}_1$ using the normal case procedure are valid for almost any type of nonnormal or unspecified distribution, provided the $\mathbf{f}_t$ and the errors are independent. This result validated the wide use of the standard factor analysis procedure. In this section, we discuss the effect of error heteroscedasticity on the asymptotic properties of the standard estimator of $\boldsymbol{\beta}_1$ obtained under the homoscedastic model.

Consider the maximum likelihood estimator $\tilde{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ where model (1.3) is fitted assuming $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$ are normal. But, suppose that the heteroscedastic model (3.1) is the true model for the data with i.i.d. $\mathbf{f}_t$. Then, under mild conditions on moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$, a result of Anderson and Amemiya (1988) holds, and

$$\sqrt{n}\,(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = \mathbf{K}\,\mathbf{l}_n + o_p(1),$$

where $\mathbf{K}$ is a fixed matrix not depending on ϵ_t and $\mathbf{l}_n$ consists of the terms

$$T_{1,ij} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(f_{it} - \mu_{fi})\epsilon_{jt}, \tag{4.1}$$

$$T_{2,lm} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_{lt}\epsilon_{mt}, \quad l \neq m. \tag{4.2}$$

Here, we denote $\mu_{fi} = E(f_{it})$. Hence, $\tilde{\boldsymbol{\beta}}_1$ is still consistent and asymptotically normal. To see the effect of error heteroscedasticity on the asymptotic inference procedure, we need to consider the form of the limiting variance-covariance matrix of (4.1) and (4.2).

Under the homoscedastic model, the limiting covariance matrix of $T_{1,ij}$ and $T_{2,lm}$ is

$$\begin{pmatrix} \phi_{ii}\psi_{jj} & 0 \\ 0 & \psi_{ll}\psi_{mm} \end{pmatrix}, \tag{4.3}$$

where ϕ_{ii} is a diagonal element of $\Phi = Var(\mathbf{f}_t)$, and $\psi_{jj} = Var(\epsilon_{jt})$. Also, all $T_{1,ij}$ and $T_{2,lm}$, $l \neq m$, are uncorrelated in the limit. The estimator of (4.3) under the homoscedastic model is

$$\begin{pmatrix} \tilde{\phi}_{ii}\tilde{\psi}_{jj} & 0 \\ 0 & \tilde{\psi}_{ll}\tilde{\psi}_{mm} \end{pmatrix}, \tag{4.4}$$

where $\tilde{\phi}_{ii}$ and $\tilde{\psi}_{jj}$ are the maximum likelihood estimators.

Under the heteroscedastic model (3.1),

$$\begin{aligned} (f_{it} - \mu_{fi})\epsilon_{jt} &= (f_{it} - \mu_{fi})g_j(\mathbf{f}_t; \boldsymbol{\alpha})\epsilon^0_{jt}, \\ \epsilon_{lt}\epsilon_{mt} &= g_l(\mathbf{f}_t; \boldsymbol{\alpha})g_m(\mathbf{f}_t; \boldsymbol{\alpha})\epsilon^0_{lt}\epsilon^0_{mt}, \end{aligned}$$

where $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ is either a polynomial or a square root of a polynomial in $\mathbf{f}_t$. Thus, for i.i.d. $\mathbf{f}_t$, with some mild conditions, the limiting covariance matrix of $T_{1,ij}$ and $T_{2,lm}$ in (4.1) and (4.2) is

$$\begin{pmatrix} E[(f_{it} - \mu_{fi})^2 g_j^2(\mathbf{f}_t; \boldsymbol{\alpha})] & 0 \\ 0 & E[g_l^2(\mathbf{f}_t; \boldsymbol{\alpha})g_m^2(\mathbf{f}_t; \boldsymbol{\alpha})] \end{pmatrix}, \tag{4.5}$$

where we used $E(\epsilon_{it}^{02}) = 1$. In this case, all $T_{1,ij}$ and $T_{2,lm}$, $l \neq m$, are also uncorrelated in the limit.

The homoscedastic estimators $\tilde{\phi}_{ii}$ and $\tilde{\psi}_{jj}$ are consistent under the heteroscedastic model (3.1) with i.i.d $\mathbf{f}_t$ for $Var(f_{it}) = E(f_{it} - \mu_{fi})^2$ and $E(g_j^2(\mathbf{f}_t; \boldsymbol{\alpha}))$ respectively. Thus the estimated covariance (4.4) is consistent for

$$\begin{pmatrix} E[(f_{it} - \mu_{fi})^2]E[g_j^2(\mathbf{f}_t; \boldsymbol{\alpha})] & 0 \\ 0 & E[g_l^2(\mathbf{f}_t; \boldsymbol{\alpha})]E[g_m^2(\mathbf{f}_t; \boldsymbol{\alpha})] \end{pmatrix}. \tag{4.6}$$

Note that (4.5) and (4.6) are different in general, and thus the asymptotic inference for $\boldsymbol{\beta}_1$ using the homoscedastic variance estimate is not valid for the heteroscedastic model (3.1). However, the difference between (4.5) and (4.6) is generally small, and can be negligible depending on the function $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ and the distribution of $\mathbf{f}_t$. Hence, we expect homoscedastic inferences for $\boldsymbol{\beta}_1$ are approximately valid asymptotically for many heteroscedastic models of the form (3.1) with $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ being a low-order polynomial.

As an example we consider the one-factor model (3.5) with $g_i(\mathbf{f}_t; \boldsymbol{\alpha}) = \alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2$. We write f_t and ϕ^2 in place of f_{it} and ϕ_{ii}. The difference between matrices (4.5) and (4.6) consists of

$$\begin{aligned} &E\left[(f_t - \mu_f)^2(\alpha_{0j} + \alpha_{1j} f_t + \alpha_{2j} f_t^2)\right] - \phi^2 E(\alpha_{0j} + \alpha_{1j} f_t + \alpha_{2j} f_t^2) \\ &\quad = E[(f_t - \mu_f)^3](\alpha_{1i} + 2\alpha_{2i}\mu_f) + \alpha_{2i}\left(E[(f_t - \mu_f)^4] - \phi^4\right), \end{aligned}$$

and

$$\begin{aligned} &E\left[(\alpha_{0l} + \alpha_{1l} f_t + \alpha_{2l} f_t^2)(\alpha_{0m} + \alpha_{1m} f_t + \alpha_{2m} f_t^2)\right] \\ &-E(\alpha_{0l} + \alpha_{1l} f_t + \alpha_{2l} f_t^2)E(\alpha_{0m} + \alpha_{1m} f_t + \alpha_{2m} f_t^2) \\ &\quad = \phi^2(\alpha_{1l} + 2\alpha_{2l}\mu_f)(\alpha_{1m} + 2\alpha_{2m}\mu_f) \\ &\qquad + E[(f_t - \mu_f)^3]\left\{\alpha_{2l}(\alpha_{1m} + 2\alpha_{2m}\mu_f) + \alpha_{2m}(\alpha_{1l} + 2\alpha_{2l}\mu_f)\right\} \\ &\qquad + \alpha_{2l}\alpha_{2m}\left(E[(f_t - \mu_f)^4] - \phi^4\right). \end{aligned}$$

The magnitude of these differences are determined by the first four factor moments and the heteroscedasticity parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. These are large when the skewness and kurtosis of $\mathbf{f}_t$ are large. If the underlying factor distribution is symmetric and has small kurtosis, e.g., normal or uniform factors, then the differences are small. Note that these differences need to be assessed in relation to the overall covariance matrix. Hence, the difference in the actual inference procedure tends to be small in many situations, and the asymptotic inference using the homoscedastic $\tilde{\boldsymbol{\beta}}_1$ may not be in serious error. But, the homoscedastic analysis does not permit any modeling of error variances, and may not be very efficient in factor score inference.

4.2 Asymptotic Properties of the Heteroscedastic Model Estimators

Let the heteroscedastic model (3.1) hold, and consider $\hat{\boldsymbol{\theta}}_{WLS}$ and $\hat{\boldsymbol{\theta}}_{PL}$ as defined in (3.10) and (3.13). If $\mathbf{f}_t$, $t = 1, 2, \ldots, n$, are i.i.d., the limiting distributions of $\hat{\boldsymbol{\theta}}_{WLS}$ and $\hat{\boldsymbol{\theta}}_{PL}$ can be derived using the standard arguments. Such results justify the use of the estimated approximate covariance matrices $\widehat{\mathbf{V}}_{WLS}$ and $\widehat{\mathbf{V}}_{PL}$ in (3.11) and (3.15) for asymptotic inferences on $\boldsymbol{\theta} = (\boldsymbol{\beta}_0', \ (vec\boldsymbol{\beta}_1)', \ \boldsymbol{\alpha}', \ \boldsymbol{\theta}_{f\epsilon}')' = (\boldsymbol{\theta}_1', \ \boldsymbol{\theta}_{f\epsilon}')'$, where $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_0', \ (vec\boldsymbol{\beta}_1)', \ \boldsymbol{\alpha}')'$. In practice, our interest is mostly in making inferences about $\boldsymbol{\theta}_1$, while the moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ are rarely of direct interest. To check the moments of $\boldsymbol{\epsilon}_t^0$, e.g., the symmetry of error or conditional normality of error given $\mathbf{f}_t$, the moments of $\boldsymbol{\epsilon}_t^0$ may be of some interest. We divide the moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ in $\boldsymbol{\theta}_{f\epsilon}$, and write

$$\boldsymbol{\theta}_{f\epsilon} = \begin{pmatrix} \boldsymbol{\theta}_\epsilon \\ \boldsymbol{\theta}_f \end{pmatrix}$$

with $\boldsymbol{\theta}_\epsilon$ consisting of the moments of $\boldsymbol{\epsilon}_t^0$, so that

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_\epsilon \\ \boldsymbol{\theta}_f \end{pmatrix}.$$

We also write $\hat{\boldsymbol{\theta}}_{PL,1}$, $\hat{\boldsymbol{\theta}}_{WLS,1}$, $\hat{\boldsymbol{\theta}}_{PL,\epsilon}$, $\hat{\boldsymbol{\theta}}_{WLS,\epsilon}$ for parts of $\hat{\boldsymbol{\theta}}_{PL}$ and $\hat{\boldsymbol{\theta}}_{WLS}$ corresponding to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_\epsilon$, and let

$$\hat{\boldsymbol{\theta}}_{PL,1\epsilon} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{PL,1} \\ \hat{\boldsymbol{\theta}}_{PL,\epsilon} \end{pmatrix}, \tag{4.7}$$

$$\hat{\boldsymbol{\theta}}_{WLS,1\epsilon} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{WLS,1} \\ \hat{\boldsymbol{\theta}}_{WLS,\epsilon} \end{pmatrix}. \tag{4.8}$$

Let

$$\widehat{\mathbf{V}}_{PL} = \begin{pmatrix} \widehat{\mathbf{V}}_{PL,1\epsilon} & \widehat{\mathbf{C}}_{PL} \\ \widehat{\mathbf{C}}'_{PL} & \widehat{\mathbf{V}}_{PL,2} \end{pmatrix}, \tag{4.9}$$

$$\widehat{\mathbf{V}}_{WLS} = \begin{pmatrix} \widehat{\mathbf{V}}_{WLS,1\epsilon} & \widehat{\mathbf{C}}_{WLS} \\ \widehat{\mathbf{C}}'_{WLS} & \widehat{\mathbf{V}}_{WLS,2} \end{pmatrix}, \tag{4.10}$$

where $\widehat{\mathbf{V}}_{PL,1\epsilon}$ and $\widehat{\mathbf{V}}_{WLS,1\epsilon}$ correspond to $\hat{\boldsymbol{\theta}}_{PL,1\epsilon}$ and $\hat{\boldsymbol{\theta}}_{WLS,1\epsilon}$. It turns out that the asymptotic inferences for $\boldsymbol{\theta}_{1\epsilon}$ using $(\hat{\boldsymbol{\theta}}_{PL,1\epsilon}, \widehat{\mathbf{V}}_{PL,1\epsilon})$ or $(\hat{\boldsymbol{\theta}}_{WLS,1\epsilon}, \widehat{\mathbf{V}}_{WLS,1\epsilon})$ are also valid for non-i.i.d. $\mathbf{f}_t$. This results holds for almost any type of $\mathbf{f}_t$. For example, the factors $\mathbf{f}_t$ can be correlated over t, fixed quantities, coming from multiple populations, or having heteroscedastic distributions. This is of practical importance, because the asymptotic inference procedures for $\boldsymbol{\theta}_{1\epsilon}$ are useful for multi-sample study, time-series analysis, longitudinal data, and non-random samples.

For notational simplicity, we first consider $\hat{\boldsymbol{\theta}}_{WLS}$. Let the true parameter value of $\boldsymbol{\theta}_1$ be denoted by $\boldsymbol{\theta}_1^0$. Under (3.1), with polynomial $g_1(\mathbf{f}_t; \boldsymbol{\alpha})$, we can write $\bar{\mathbf{A}}$ in (3.8) as

$$\bar{\mathbf{A}} = \mathbf{b}(\boldsymbol{\theta}_1^0) + \mathbf{B}_1(\boldsymbol{\theta}_1^0)\mathbf{h}_{1,n} + \mathbf{B}_2(\boldsymbol{\theta}_1^0)\mathbf{h}_{2,n}, \tag{4.11}$$

where the vector $\mathbf{b}(\boldsymbol{\theta}_1^0)$ and the matrices $\mathbf{B}_1(\boldsymbol{\theta}_1^0)$ and $\mathbf{B}_2(\boldsymbol{\theta}_1^0)$ do not involve $\mathbf{f}_t$ or ϵ_t, $\mathbf{h}_{1,n}$ is the vector (free of $\boldsymbol{\theta}_1^0$) consisting of sample moments of $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{kt})'$ of the form

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{r_1} f_{2t}^{r_2} \ldots f_{kt}^{r_k}, \tag{4.12}$$

and $\mathbf{h}_{2,n}$ is the vector (free of $\boldsymbol{\theta}_1^0$) consisting of sample cross-products between powers of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0 = (\epsilon_{1t}^0, \epsilon_{2t}^0, \ldots, \epsilon_{pt}^0)'$ of the form

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{m_1} f_{2t}^{m_2} \ldots f_{kt}^{m_k} \epsilon_{1t}^{0q_1} \epsilon_{2t}^{0q_2} \ldots \epsilon_{pt}^{0q_p}. \tag{4.13}$$

Note that r_i, m_i and q_i appearing in (4.12) and (4.13) are determined by the function $g_i(\mathbf{f}_t; \boldsymbol{\alpha})$ and the choice of $\mathbf{U}_t$ in (3.2). We assume that $\mathbf{h}_{1,n}$ and $\mathbf{h}_{2,n}$ list distinct sample moments without duplication.

For $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$, we assume

(i) $\mathbf{f}_t$, ϵ_{1t}^0, ϵ_{2t}^0, ..., ϵ_{pt}^0 are independent,

(ii) for each $i = 1, 2, \ldots, p$, ϵ_{it}^0's are i.i.d.

(iii) for every term of form ϵ_{it}^{0q} appearing in (4.13) in $\mathbf{h}_{2,n}$, $E(\epsilon_{it}^{0^{2q}}) < \infty$,

(iv) $\mathbf{h}_{1,n}$ in (4.11-4.12) satisfies the condition that, as $n \to \infty$,

$$\mathbf{h}_{1,n} \to \mathbf{h}_{1,0} \quad a.s.. \tag{4.14}$$

For every term in $\mathbf{h}_{1,n}$ of form (4.12),

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{2r_1} f_{2t}^{2r_2} \ldots f_{kt}^{2r_k}, \tag{4.15}$$

converges almost surely as $n \to \infty$. Also, every term of form (4.13) appearing in $\mathbf{h}_{2,n}$ satisfies that, as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{m_1} f_{2t}^{m_2} \ldots f_{kt}^{m_k} \to \eta_{m_1,m_2,\ldots,m_k}, \quad a.s., \tag{4.16}$$

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{2m_1} f_{2t}^{2m_2} \ldots f_{kt}^{2m_k} \to \tau_{m_1,m_2,\ldots,m_k}, \quad a.s.. \tag{4.17}$$

Note that a large class of random or fixed $\mathbf{f}_t$ satisfy (iv), and that $\mathbf{f}_t$'s do not have to be i.i.d. or homoscedastic, and may be a stationary process over t. Recall that $\boldsymbol{\zeta}(\boldsymbol{\theta})$ used in defining $\hat{\boldsymbol{\theta}}_{WLS}$ in (3.10) was $E(\bar{\mathbf{A}})$ obtained under i.i.d. $\mathbf{f}_t$, and that $\boldsymbol{\theta}_{f\epsilon}$ in $\boldsymbol{\theta}$ consisted of all moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ appearing in $E(\bar{\mathbf{A}})$ for i.i.d. $\mathbf{f}_t$. Using (4.11), $\boldsymbol{\theta}_{f\epsilon}$ lists all distinct moments of $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$ appearing in

$$E\left\{\begin{pmatrix} \mathbf{h}_{1,n} \\ \mathbf{h}_{2,n} \end{pmatrix}\right\},$$

where the expectation is taken as if $\mathbf{f}_t$'s are i.i.d.. Under the distribution of $\boldsymbol{\epsilon}_t^0$ in (ii), the true value of $\boldsymbol{\theta}_\epsilon$ is denoted by $\boldsymbol{\theta}_\epsilon^0$. Let $\boldsymbol{\theta}_{1\epsilon}^0 = (\boldsymbol{\theta}_1^{0\prime},\ \boldsymbol{\theta}_\epsilon^{0\prime})'$. Note that $\mathbf{f}_t$ may not be i.i.d. under (iv), and there may not be a well-defined "true value" of $\boldsymbol{\theta}_f$. But, for every element of $\boldsymbol{\theta}_f$, there is a corresponding sample moments of n $\mathbf{f}_t$'s of form (4.16) either included in (4.16) or $\mathbf{h}_{1,n}$. Let $\boldsymbol{\theta}_f(n)$ denote the vector of such distinct sample moments of $\mathbf{f}_t$ appearing in (4.16) and $\mathbf{h}_{1,n}$. Also, under (iv), there exist $\boldsymbol{\theta}_{f,\infty}$ such that, as $n \to \infty$,

$$\boldsymbol{\theta}_f(n) \to \boldsymbol{\theta}_{f,\infty}\ , \quad a.s..$$

Using these two quantities, we define two type of "true values" of $\boldsymbol{\theta}$,

$$\begin{aligned} \boldsymbol{\theta}(n) &= \begin{pmatrix} \boldsymbol{\theta}_{1\epsilon}^0 \\ \boldsymbol{\theta}_f(n) \end{pmatrix}, \\ \boldsymbol{\theta}_\infty &= \begin{pmatrix} \boldsymbol{\theta}_{1\epsilon}^0 \\ \boldsymbol{\theta}_{f,\infty} \end{pmatrix}. \end{aligned} \tag{4.18}$$

Let Θ be the parameter space over which (3.10) is minimized. For the identification condition, we assume

(v) $\boldsymbol{\theta}_\infty$ is an interior point of Θ, the matrix

$$\Delta^* = \frac{\partial \boldsymbol{\zeta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\infty}$$

has full column rank, and for any $\gamma_1 > 0$, $\exists$ a $\gamma_2 > 0$ satisfying

$$|\boldsymbol{\theta} - \boldsymbol{\theta}_\infty| > \gamma_1 \Rightarrow |\boldsymbol{\zeta}(\boldsymbol{\theta}) - \boldsymbol{\zeta}(\boldsymbol{\theta}_\infty)| > \gamma_2.$$

Then, we have the following result.

Theorem 1 *Let model (3.1) hold with polynomial $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$, and let assumptions (i)-(v) hold. Then, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{WLS,1\epsilon} - \boldsymbol{\theta}^0_{1\epsilon}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Upsilon_{1\epsilon}), \tag{4.19}$$

$$n\widehat{\mathbf{V}}_{WLS,1\epsilon} \xrightarrow{p} \Upsilon_{1\epsilon}. \tag{4.20}$$

where $\hat{\boldsymbol{\theta}}_{WLS,1\epsilon}$ and $\widehat{\mathbf{V}}_{WLS,1\epsilon}$ are defined in (4.8) and (4.10).

Proof By (4.15) and (4.17) in (iv) and by (iii), $\hat{\Pi}$ in (3.10) satisfies

$$n\hat{\Pi} \longrightarrow \Pi_0, \ a.s..$$

Also,

$$\bar{\mathbf{A}} \longrightarrow \boldsymbol{\zeta}(\boldsymbol{\theta}_\infty), \ a.s..$$

Note that the singularity in Π_0 corresponds to the redundancy in $\boldsymbol{\zeta}(\boldsymbol{\theta})$. Thus the identification condition (v) implies that

$$\hat{\boldsymbol{\theta}}_{WLS} \longrightarrow \boldsymbol{\theta}_\infty, \ a.s.. \tag{4.21}$$

Since $\boldsymbol{\theta}_\infty$ is an interior point of Θ, with probability approaching one,

$$\left.\frac{\partial \boldsymbol{\zeta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{WLS}} (n\hat{\Pi})^+ \left[\bar{\mathbf{A}} - \boldsymbol{\zeta}(\hat{\boldsymbol{\theta}}_{WLS})\right] = \mathbf{0}. \tag{4.22}$$

Expanding (4.22) for $\hat{\boldsymbol{\theta}}_{WLS}$ around $\boldsymbol{\theta}(n)$, and using (4.21), we have

$$\hat{\boldsymbol{\theta}}_{WLS} - \boldsymbol{\theta}(n) = \left(\Delta^{*\prime}\Pi_0^+\Delta^*\right)^{-1} \Delta^{*\prime}\Pi_0^+\mathbf{A}_n^* + \mathbf{R}, \tag{4.23}$$

where

$$\mathbf{A}_n^* = \bar{\mathbf{A}} - \boldsymbol{\zeta}(\boldsymbol{\theta}(n)),$$

and $\mathbf{R}$ is of smaller order in probability than $\mathbf{A}_n^*$. By (4.11), (4.12), (4.13) and the definition of $\boldsymbol{\theta}(n)$ in (4.18),

$$\begin{aligned}\boldsymbol{\zeta}(\boldsymbol{\theta}(n)) &= E(\bar{\mathbf{A}}|\mathbf{f}_t's) \\ &= \mathbf{b}(\boldsymbol{\theta}_1^0) + \mathbf{B}_1(\boldsymbol{\theta}_1^0)\mathbf{h}_{1,n} + \mathbf{B}_2(\boldsymbol{\theta}_1^0)E(\mathbf{h}_{2,n}|\mathbf{f}_t's),\end{aligned}$$

where $\mathbf{h}_{1,n}$ is a part of $\boldsymbol{\theta}_f(n)$ and the elements of $E(\mathbf{h}_{2,n}|\mathbf{f}_t's)$ have the form

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{m_1} f_{2t}^{m_2} \ldots f_{kt}^{m_k} E\left(\epsilon_{1t}^{0q_1}\epsilon_{2t}^{0q_2}\ldots\epsilon_{pt}^{0q_p}\right)$$

with the sample moments of $\mathbf{f}_t$ being a part of $\boldsymbol{\theta}_f(n)$. Thus,

$$\mathbf{A}_n^* = \mathbf{B}_2(\boldsymbol{\theta}_1^0)\mathbf{H}_{2,n}\ , \tag{4.24}$$

where the elements of $\mathbf{H}_{2,n}$ have the form

$$\frac{1}{n}\sum_{t=1}^{n} f_{1t}^{m_1} f_{2t}^{m_2} \ldots f_{kt}^{m_k} \left[\epsilon_{1t}^{0q_1}\epsilon_{2t}^{0q_2}\ldots\epsilon_{pt}^{0q_p} - E\left(\epsilon_{1t}^{0q_1}\epsilon_{2t}^{0q_2}\ldots\epsilon_{pt}^{0q_p}\right)\right]. \tag{4.25}$$

To see the limiting distribution of $\mathbf{A}_n^*$, first we condition on the $\mathbf{f}_t$'s satisfying (4.17) in (iv). Then, by (i), (iii) and (4.17), a version of central limit theorem (see, for example, Lemma 1 of Amemiya and Fuller (1984)) applies to $\mathbf{H}_{2,n}$ with a typical element (4.25), and

$$\sqrt{n}\,\mathbf{H}_{2,n} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_H), \tag{4.26}$$

where $\mathbf{V}_H$ depends only on the moment of $\boldsymbol{\epsilon}_t^0$ as in (iii) and the limits $\tau_{m_1,m_2,\ldots,m_k}$ in (4.17). Since $\mathbf{f}_t$'s satisfy (4.17) with probability one under (iv), (4.26) holds unconditionally by the dominated convergence theorem. Hence, by (4.23), (4.24) and (4.26),

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_{WLS} - \boldsymbol{\theta}(n)) \xrightarrow{\mathcal{L}} N(\mathbf{0},\ \boldsymbol{\Upsilon}), \tag{4.27}$$

where

$$\Upsilon = (\Delta^{*\prime}\Pi_0^{+}\Delta^{*})^{-1}\,\Delta^{*\prime}\Pi_0^{+}\mathbf{V}_H\Pi_0^{+}\Delta^{*}\,(\Delta^{*\prime}\Pi_0^{+}\Delta^{*})^{-1}.$$

By (4.18), the first part of $\boldsymbol{\theta}(n)$ is $\boldsymbol{\theta}_{1\epsilon}$. Thus the first part of (4.27) is (4.19), where $\Upsilon_{1\epsilon}$ is the corresponding part of Υ. As noted in (4.26), $\Upsilon_{1\epsilon}$ depends only on the moments of ϵ_t^0 and $\tau_{m_1,m_2,\ldots,m_k}$. Note also that $\widehat{\mathbf{V}}_{WLS,1\epsilon}$ uses the moment-based estimator $\widehat{\Pi}$ in (3.9). Hence, if $\mathbf{f}_t$'s are i.i.d, then (4.20) holds because of convergence of the sample moments to true moments. But, by (iii) and (iv), such a convergence also holds for $\mathbf{f}_t$ satisfying (iv). Thus (4.20) follows. ■

It can be shown that, under essentially the same conditions, $\hat{\boldsymbol{\theta}}_{PL,1\epsilon}$ and $\widehat{\mathbf{V}}_{PL,1\epsilon}$ satisfy the results corresponding to (4.19) and (4.20). The precise statement and proof are omitted for brevity. Hence, the asymptotic inference for $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_0', \ (vec\,\boldsymbol{\beta}_1)', \ \boldsymbol{\alpha}')'$ (and $\boldsymbol{\theta}_\epsilon$ if necessary) can be carried out correctly using $(\hat{\boldsymbol{\theta}}_{PL,1\epsilon}, \widehat{\mathbf{V}}_{PL,1\epsilon})$ or $(\hat{\boldsymbol{\theta}}_{WLS,1\epsilon}, \widehat{\mathbf{V}}_{WLS,1\epsilon})$ for a wide class of $\mathbf{f}_t$'s. Note also that the distributional form of ϵ_t^0 is also unspecified. Thus, our inference procedures are asymptotically distribution-free methods that are applicable for virtually any type of $\mathbf{f}_t$ and ϵ_t^0 appearing in practice.

Up to this point, we assumed that $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$ in (3.1) is a polynomial in $\mathbf{f}_t$. An alternative specification of the heteroscedasticity assumes that $g_i^2(\mathbf{f}_t;\boldsymbol{\alpha})$ is a polynomial in $\mathbf{f}_t$, i.e., that $g_i(\mathbf{f}_t;\boldsymbol{\alpha})$ is the square root of a polynomial. For such a model, the results identical to (4.19) and (4.20) for $\hat{\boldsymbol{\theta}}_{PL,1\epsilon}$ and $\hat{\boldsymbol{\theta}}_{WLS,1\epsilon}$ hold, provided that the choice of $\mathbf{Z}_{a,t}$ in (3.2) is made to ensure that no square root appears in $\boldsymbol{\mu}_a$, $\Sigma_a(\boldsymbol{\theta})$ and $\boldsymbol{\zeta}(\boldsymbol{\theta})$ in (3.3) and (3.12). One such case is the model (3.5), when only $W_{ij,t}$ are included in $\mathbf{Z}_{a,t}$ (but not Y_{it}). This model was considered in Chapter 3 and is used in the simulation study in the next chapter.

CHAPTER 5. SIMULATION STUDY

In the previous chapters, we presented a moment-based approach to fitting a factor model with heteroscedastic errors, and discussed estimation of the true factor value. The following sections present the results of two simulation studies. The purpose of the first simulation study is to assess and compare the proposed model-fitting procedures, and the results are discussed in Section 5.1. Since a strength of the proposed procedures is the absence of distributional assumptions for $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$, various distributional forms were considered for $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t^0$. Different sample sizes were also considered. The second simulation study was conducted to study the behavior of the factor score estimators and their estimated standard errors. These results are presented in Section 5.2. The computer programs for both simulation studies were written in SAS, and includes the use of SAS/IML, SAS MACROS, and PROC CALIS. Some overall recommendations on the use of the procedures are made in Section 5.3.

5.1 Simulation Study I

We first compare the homoscedastic and heteroscedastic (PL and WLS) estimators for $\boldsymbol{\beta}_1$ and $\boldsymbol{\alpha}$. It will be shown that the PL estimators have more desirable properties in finite samples than the WLS. The performance of the PL estimators under non-i.i.d. factors was also studied.

Throughout, we assumed the following one-factor model with $p = 4$,

$$\mathbf{Z}_t = \begin{pmatrix} \boldsymbol{\beta}_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1 \\ 1 \end{pmatrix} f_t + \boldsymbol{\epsilon}_t, \tag{5.1}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t},\ \epsilon_{2t},\ \epsilon_{3t},\ \epsilon_{4t})'$, and true $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ values given by

$$\begin{aligned} \boldsymbol{\beta}_0 &= (\beta_{01},\ \beta_{02},\ \beta_{03})' = (1,\ 2,\ -3)', \\ \boldsymbol{\beta}_1 &= (\beta_{11},\ \beta_{21},\ \beta_{31})' = (6,\ 5,\ 4)'. \end{aligned} \tag{5.2}$$

The true error variance structure is

$$\epsilon_{it} = \sqrt{\alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2}\, \epsilon_{it}^0. \tag{5.3}$$

For the distributions of f_t and $\boldsymbol{\epsilon}_t$, we considered three different cases

A. $f_t \sim N(1.7,\ 0.4)$, $\epsilon_{it}^0 \sim N(0,\ 1)$, $i = 1, 2, 3, 4$,

B. $f_t \sim Uniform(0,\ 3)$, $\epsilon_{it}^0 \sim Uniform(-2,\ 2) \times \sqrt{\frac{12}{16}}$, $i = 1, 2, 3, 4$,

C. $f_t \sim \frac{\chi_{15}^2 - 15}{\sqrt{30}} \times \sqrt{0.75} + 1.5$, $\epsilon_{it}^0 \sim N(0,\ 1)$, $i = 1, 2, 3, 4$.

For the true $\boldsymbol{\alpha}$ values, we considered two sets; homoscedastic and mixed heteroscedastic models. For the mixed heteroscedastic model, the $\boldsymbol{\alpha}$ values are

$$\begin{aligned} \boldsymbol{\alpha}_0 &= (\alpha_{01},\ \alpha_{02},\ \alpha_{03},\ \alpha_{04})' = (3,\ 2,\ 1.5,\ 1.5)', \\ \boldsymbol{\alpha}_1 &= (\alpha_{11},\ \alpha_{12},\ \alpha_{13},\ \alpha_{14})' = (1,\ 0,\ 1,\ 0)', \\ \boldsymbol{\alpha}_2 &= (\alpha_{21},\ \alpha_{22},\ \alpha_{23},\ \alpha_{24})' = (2,\ 2,\ 0,\ 0)'. \end{aligned} \tag{5.4}$$

For the homoscedastic models, $\boldsymbol{\alpha}_0$ in (5.4) was used with $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{0}$, so that the true error variance is given by $\Psi_0 = diag(3,\ \ 2,\ \ 1.5,\ \ 1.5)$. Thus, there are six combinations depending on the three distributional forms and two error models.

For each of the six cases, 1000 samples each of size 1000, and 1000 samples each of size 300 were generated. For each sample, we applied three methods of estimation.

The first two are the pseudo-likelihood and weighted least squares methods proposed in Chapter 3, using all possible $W_{ij,t}=(Z_{it}-\bar{Z}_i)(Z_{jt}-\bar{Z}_j)$ as additional variables. For these two methods, model (5.1) with (5.3) was fitted to each dataset. The third method is the standard (maximum likelihood) factor modeling approach which disregards any possible heteroscedasticity in the model, and fits (5.1) with error variance assumed to be a fixed diagonal matrix. The results from the first two methods will be denoted by PL and WLS while that for the standard approach will be denoted by HOM. For each of the $\boldsymbol{\beta}_1$ parameters, the 95% confidence intervals (c.i.) is given by

$$\hat{\beta}_{i1} \pm 1.96\sqrt{\widehat{Var}(\hat{\beta}_{1i})},$$

where $\hat{\beta_{1i}}$ is the estimate of β_{1i} from any one of the three estimation methods, and $\widehat{Var}(\hat{\beta}_{1i})$ is its corresponding variance estimate using the formulae (3.11) and (3.15) for the WLS and PL methods. The 95% c.i.'s for each of the $\boldsymbol{\alpha}$ parameters are constructed in a similar fashion.

We first summarize the results for the homoscedastic true model with the three different distributional forms. Figure 5.1 gives the boxplot of the estimation errors for β_{12}, (β_{12} estimates - true value of β_{12}) for Case B with sample size 1000. The boxplots for the other β_{1i} estimates, other distributional cases, and two sample sizes are similar. The boxplots suggest that all three methods give comparably good estimates for β_{1i} across the factor and error distributions considered. This result is also reflected by their mean squared error (MSE) values in Table 5.1. Table 5.1 gives the MSE of β_{1i} estimators by the three methods. For sample size 300, the PL estimators have smaller MSE than the WLS estimators for all cases. When sample size increased to 1000, their MSE values became very similar. Overall, PL and WLS MSE values were slightly larger than those from the HOM approach. But for these homoscedastic cases, the efficiency loss of PL and WLS fitting many more parameters than HOM is small, especially in large samples.

Table 5.2 gives the coverage probabilities (the percentage of samples with the true

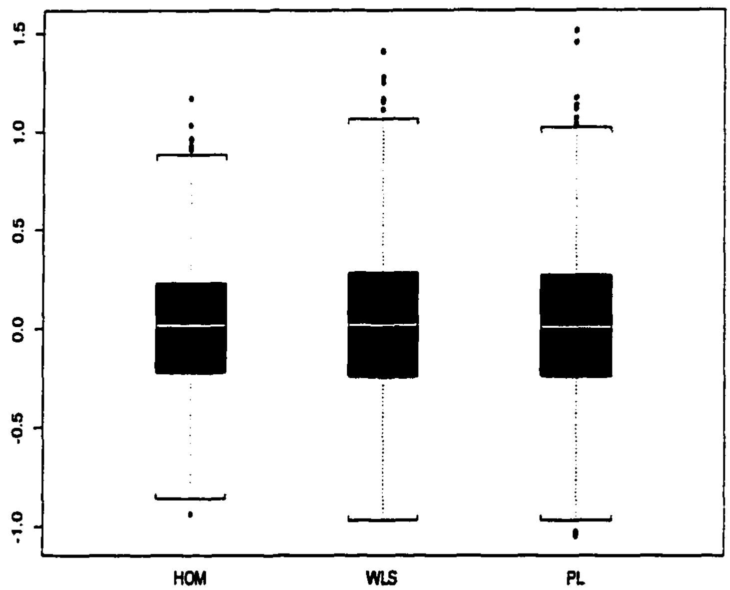

Figure 5.1 Box plots of β_{12} estimates
(homoscedastic data, Case B, n=1000)

value in the interval) of the 95% c.i.'s for β_{1i} using the three methods. The HOM and PL approaches give very accurate c.i.'s even with sample size 300. For the WLS approach, the coverage probabilities are considerably smaller than the nominal value when sample size is 300. Though the WLS results improved when the sample size was increased to 1000, it is still not as good as those of PL or HOM.

Figures 5.2 and 5.3 show the boxplots of the PL and WLS estimation errors for α_{12} and α_{22} under the homoscedastic model with Case B when sample size is 1000. The boxplots for the other α_{1i} and α_{2i} estimates, other distributional cases, and two sample sizes are similar. The values are closely clustered about zero indicating that the PL and WLS approaches can estimate the zero heteroscedasticity parameters α_{1i} and α_{2i} with small variability. Figure 5.4 plots the MSE of the PL and WLS estimators for all the α_{ij}. The MSE are very similar for the two methods, and are quite small overall.

Table 5.1 Mean squared error of β_{1i} estimators (homoscedastic data)

	β_{11}	β_{12}	β_{13}
Sample size = 300			
Normal f_t, normal ϵ_t			
PL	0.742	0.522	0.340
WLS	1.004	0.716	0.464
HOM	0.560	0.401	0.258
Uniform factor, uniform error			
PL	0.312	0.221	0.141
WLS	0.493	0.348	0.220
HOM	0.268	0.190	0.121
χ^2 factor, normal error			
PL	0.406	0.286	0.182
WLS	0.539	0.366	0.261
HOM	0.275	0.194	0.126
Sample size = 1000			
Normal f_t, normal ϵ_t			
PL	0.206	0.143	0.092
WLS	0.205	0.140	0.091
HOM	0.166	0.115	0.074
Uniform factor, uniform error			
PL	0.084	0.060	0.039
WLS	0.087	0.062	0.039
HOM	0.074	0.053	0.034
χ^2 factor, normal error			
PL	0.114	0.079	0.051
WLS	0.098	0.069	0.044
HOM	0.078	0.055	0.036

Table 5.2 Coverage probabilities of 95% c.i. for β_{1i} (homoscedastic data)

	β_{11}	β_{12}	β_{13}
	Sample size=300		
Normal f_t, normal ϵ_t			
PL	92.8	92.1	92.8
WLS	85.5	84.8	83.5
HOM	94.0	93.9	93.0
Uniform f_t, uniform ϵ_t			
PL	95.2	94.5	94.5
WLS	87.7	86.8	87.9
HOM	94.8	94.8	94.5
χ^2 f_t, normal ϵ_t			
PL	91.1	91.9	91.1
WLS	83.1	81.6	82.8
HOM	94.8	94.7	94.5
	Sample size=1000		
Normal f_t, normal ϵ_t			
PL	96.2	95.7	96.0
WLS	91.0	91.8	92.6
HOM	94.6	93.9	94.4
Uniform f_t, uniform ϵ_t			
PL	94.8	94.2	95.4
WLS	93.7	92.6	93.9
HOM	95.5	95.0	95.7
χ^2 f_t, normal ϵ_t			
PL	95.3	95.4	95.2
WLS	91.2	89.9	90.8
HOM	95.1	95.0	94.2

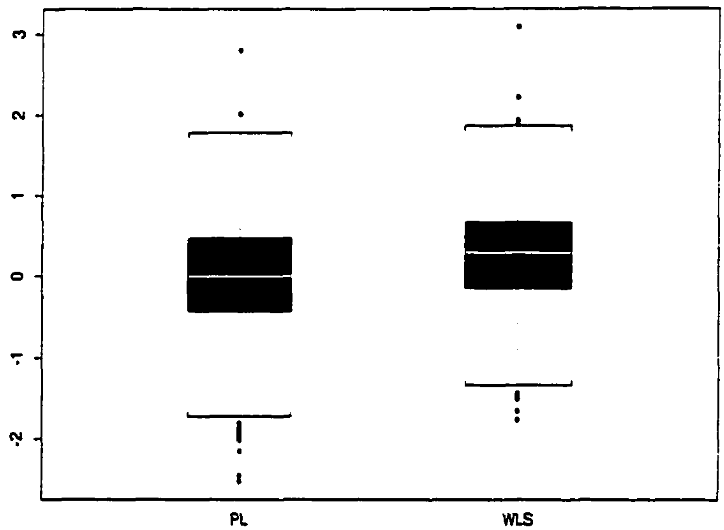

Figure 5.2 Box plots of α_{12} estimates
(homoscedastic data, Case B, n=1000)

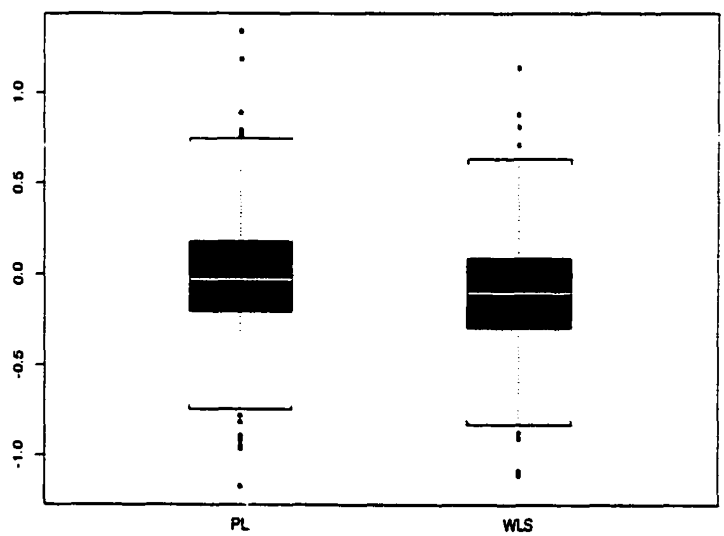

Figure 5.3 Box plots of α_{22} estimates
(homoscedastic data, Case B, n=1000)

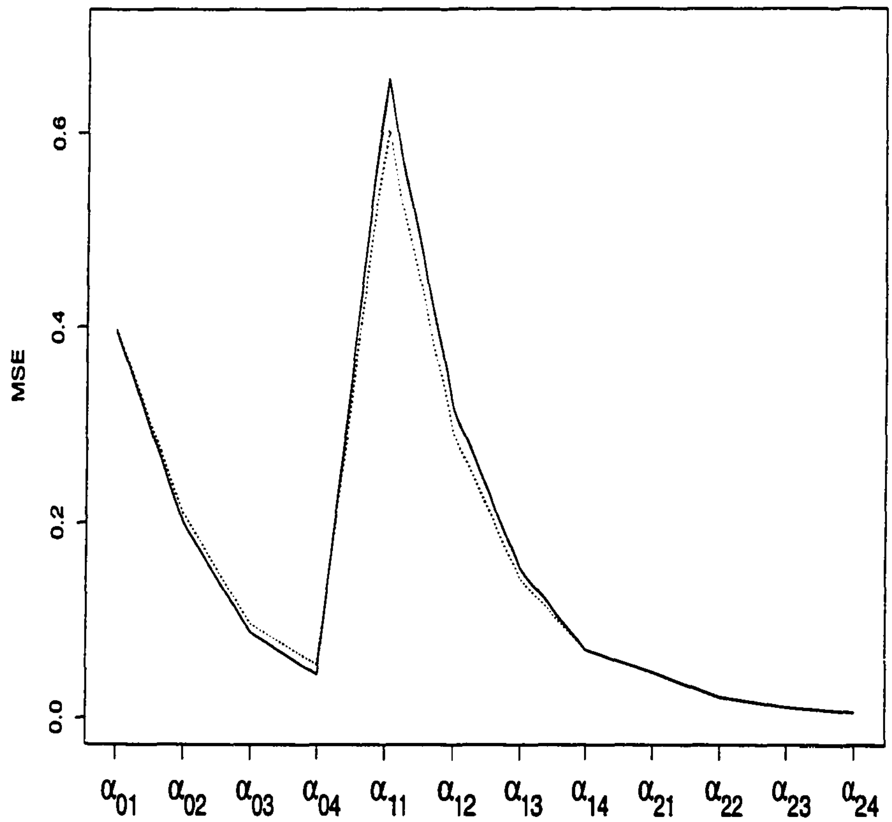

Figure 5.4 Plot of MSE of α estimates
(homoscedastic data, Case B, n=1000)
—— PL .... WLS

Table 5.3 gives the coverage probabilities of the 95% c.i.'s for α_{1i} and α_{2i} under the homoscedastic model. Since the true values of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are $\mathbf{0}$, the coverage probabilities in Table 5.3 are 100(1 - type I error) of the nominal 5% level 2-sided test for $\alpha_{1i} = 0$ and $\alpha_{2i} = 0$. The results indicate that for the PL method, the type I error is very close to the nominal level of 5% for all cases and sample sizes. For WLS, the average type I error is 10.5% for sample size 300, but became comparable to those of PL for sample size 1000.

Table 5.3 Coverage probabilities of 95 % c.i. for α_{1i} and α_{2i} (homoscedastic data)

	Normal f_t, normal ϵ_t		Uniform f_t, uniform ϵ_t		χ^2 f_t, normal ϵ_t	
	PL	WLS	PL	WLS	PL	WLS
			Sample size = 300			
α_{11}	95.5	89.4	95.8	89.9	95.6	91.2
α_{12}	95.5	89.6	96.6	92.6	96.2	92.6
α_{13}	95.9	89.1	96.1	90.0	96.5	93.2
α_{14}	95.4	85.5	94.5	89.9	97.1	91.0
α_{21}	95.6	89.0	95.7	89.7	93.6	88.9
α_{22}	95.0	87.7	96.7	91.5	94.7	89.4
α_{23}	96.0	89.7	95.7	89.3	94.7	90.6
α_{24}	95.2	83.7	93.9	88.7	94.1	86.7
			Sample size = 1000			
α_{11}	94.4	91.7	95.4	94.1	96.5	95.0
α_{12}	95.4	91.8	95.8	95.0	95.2	95.4
α_{13}	95.4	90.7	94.9	93.4	95.5	95.2
α_{14}	94.1	88.4	96.2	93.6	97.4	95.6
α_{21}	94.7	91.1	95.0	93.6	94.6	92.8
α_{22}	95.9	91.4	95.3	94.4	94.3	93.9
α_{23}	95.4	91.6	94.7	94.1	93.4	92.5
α_{24}	94.8	87.4	95.3	92.8	94.2	90.7

Now, the results for the heteroscedastic model with the three distributional cases are summarized. Figure 5.5 gives the boxplot of the estimation errors for β_{12} for Case B when sample size is 1000. The boxplots for the other β_{1i} estimates and other cases are similar. Table 5.4 gives the MSE of the β_{1i} estimators by the three methods. The MSE of PL estimators of β_{1i} were slightly smaller than the MSE of WLS estimators for sample size 300. When sample size was 1000, the MSE of WLS estimators were slightly smaller than the MSE of PL estimators. Overall, the PL and WLS approaches yielded comparably good β_{1i} estimates.

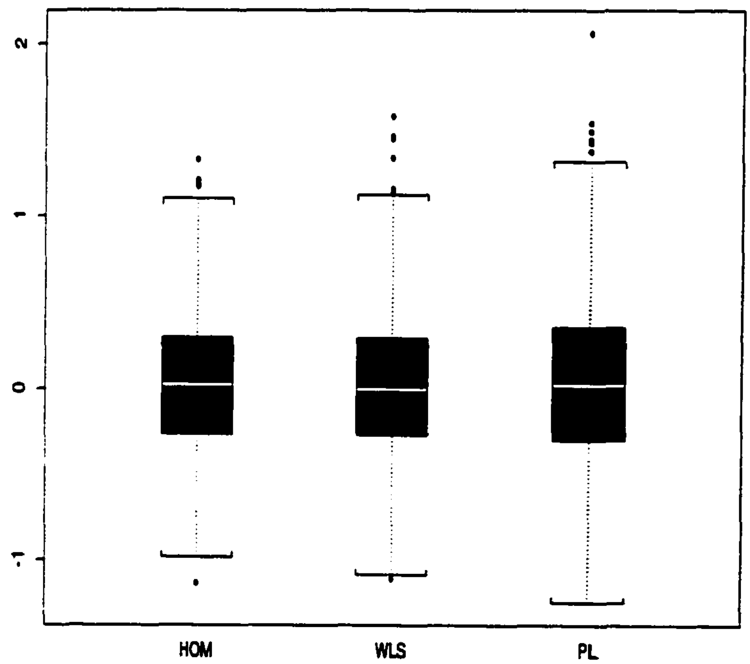

Figure 5.5 Box plots of β_{12} estimates
(heteroscedastic data, Case B, n=1000)

Table 5.4 Mean Squared Error of β_{1i} estimators (heteroscedastic data)

	β_{11}	β_{12}	β_{13}
Sample size = 300			
Normal f_t, normal ϵ_t			
PL	1.118	0.807	0.495
WLS	1.295	0.960	0.578
HOM	0.792	0.594	0.342
Uniform factor, uniform ϵ_t			
PL	0.414	0.312	0.190
WLS	0.560	0.399	0.239
HOM	0.337	0.254	0.147
χ^2 factor, normal ϵ_t			
PL	0.789	0.580	0.268
WLS	0.606	0.426	0.257
HOM	0.381	0.286	0.153
Sample size = 1000			
Normal f_t, normal ϵ_t			
PL	0.309	0.225	0.130
WLS	0.257	0.175	0.112
HOM	0.230	0.167	0.098
Uniform factor, uniform ϵ_t			
PL	0.113	0.082	0.051
WLS	0.105	0.075	0.045
HOM	0.096	0.069	0.041
χ^2 factor, normal ϵ_t			
PL	0.238	0.169	0.083
WLS	0.107	0.079	0.044
HOM	0.123	0.089	0.053

Table 5.5 gives the coverage probabilities of the nominal 95% c.i.'s for β_{1i}. The PL c.i. has coverage close to the desired 95% level. But, the WLS c.i. is not as accurate, especially with sample size 300 when the average coverage probability is only 84.7% compared to the value of 92.7% for PL. When sample size is 1000, the average coverage probability is 95.7 % for the PL, and is 91.7% for the WLS. In Chapter 4, we showed that the standard error of the standard HOM estimator is not largely affected by error heteroscedasticity. In this simulation study, with the small true $\boldsymbol{\alpha}$ values and the small third and fourth factor moments, the coverage probability of the HOM c.i. for β_{1i} was still very close to the nominal level. Note that Cases A and B have symmetric factor distributions, and that Case B has low kurtosis.

We found that among the α_{1i}'s, the spread of the estimates is largest for α_{11} and decreases for $i = 2, 3, 4$, with the estimates for α_{14} being the most closely clustered about the true value. The same pattern can also be found in the boxplots of α_{2i} estimates. Since the variances of ϵ_{1t} and ϵ_{4t} represent the extremes in severity of heteroscedasticity in this study, the results for their corresponding α_{ij} are chosen for presentation here. Figure 5.6 shows the boxplots of the estimation errors for α_{11} and α_{14} for Case B and sample size 1000. For a clearer presentation of the results, three outliers with absolute values in excess of 50 were removed from the boxplot of PL α_{11} estimates. Figure 5.7 covers α_{21} and α_{24} for the same case. The plots are similar for the other cases. Although the MSE is similar for PL and WLS, the large bias in WLS makes the WLS less attractive in practice. Table 5.6 gives the coverage probabilities of the 95% c.i.'s for α_{1i} and α_{1i}. For both sample sizes, the PL c.i. is reasonable for all parameters, while the WLS c.i. is less than satisfactory for some parameters. Table 5.7 shows the power (percentage of rejection) of the 2-sided test for $H_0 : \alpha_{1i} = 0$, $i = 1, 3$, and $H_0 : \alpha_{2j} = 0$, $j = 1, 2$, when sample size is 300. The power is quite high for both PL and WLS tests, but the PL power is consistently higher than the WLS power across all cases. The same pattern was found for sample size 1000.

Table 5.5 Coverage Probabilities of 95% c.i. for β_{1i} (heteroscedastic data)

	β_{11}	β_{12}	β_{13}
Sample size = 300			
Normal f_t, normal ϵ_t			
PL	93.2	92.2	93.6
WLS	84.3	83.2	84.2
HOM	93.2	93.6	93.5
Uniform factor, uniform ϵ_t			
PL	96.0	94.8	94.3
WLS	86.2	87.6	86.5
HOM	95.3	95.5	95.7
χ^2 factor, normal ϵ_t			
PL	89.2	90.2	90.8
WLS	83.3	83.5	83.3
HOM	95.4	95.3	94.5
Sample size = 1000			
Normal f_t, normal ϵ_t			
PL	96.2	95.0	95.7
WLS	91.5	90.9	91.7
HOM	94.1	92.9	94.4
Uniform factor, uniform ϵ_t			
PL	96.1	96.2	96.2
WLS	93.1	92.6	93.3
HOM	95.3	95.5	95.7
χ^2 factor, normal ϵ_t			
PL	94.9	96.4	94.8
WLS	90.5	90.6	91.0
HOM	93.0	92.9	94.3

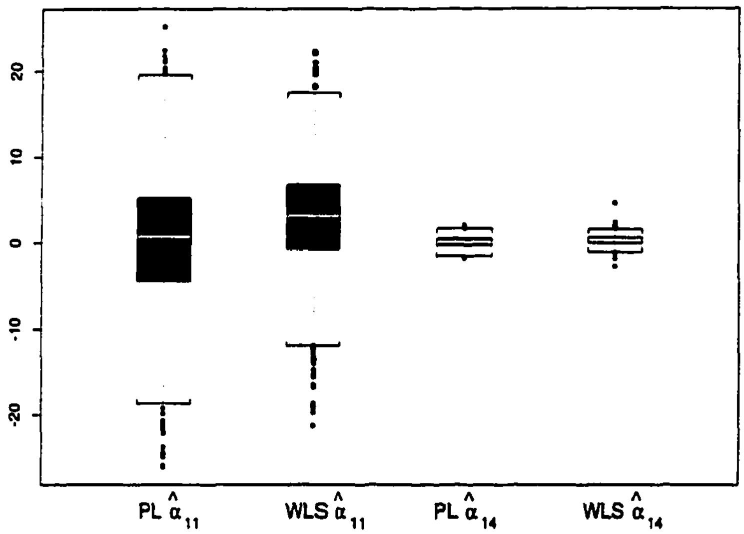

Figure 5.6 Box plots of α_{11} and α_{14} estimates
(heteroscedastic data, Case B, n=1000)

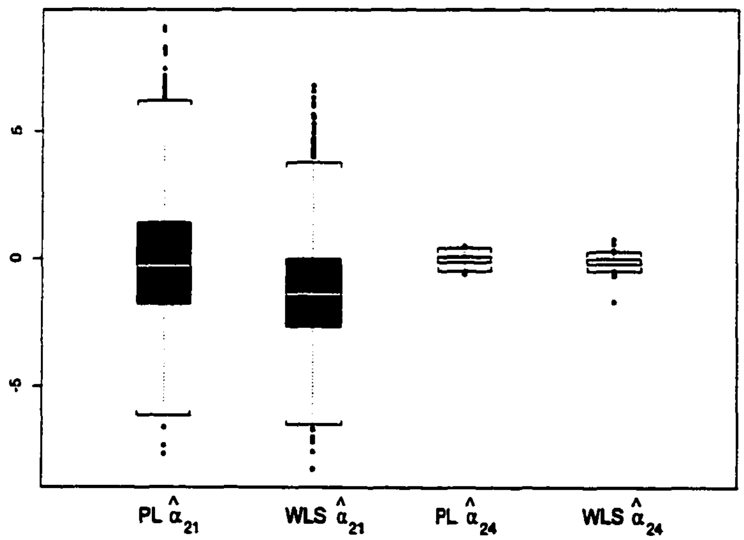

Figure 5.7 Box plots of α_{21} and α_{24} estimates
(heterocedastic data, Case B, n=1000)

Table 5.6 Coverage probabilities of 95% c.i. for α_{1i} and α_{2i} (heteroscedatic data)

	Normal f_t, normal ϵ_t		Uniform f_t, uniform ϵ_t		χ^2 f_t, normal ϵ_t	
	PL	WLS	PL	WLS	PL	WLS
			Sample size = 300			
α_{11}	94.3	85.5	94.6	86.8	88.7	85.4
α_{12}	92.4	82.4	95.0	85.2	84.7	80.0
α_{13}	97.7	92.8	95.9	93.0	93.5	92.9
α_{14}	94.6	87.8	94.3	91.6	95.1	91.6
α_{21}	88.7	72.6	92.6	80.0	77.2	68.0
α_{22}	86.8	71.0	92.9	77.7	77.6	63.8
α_{23}	94.6	85.8	94.0	88.2	88.7	87.5
α_{24}	93.3	83.2	94.4	87.1	92.0	88.8
			Sample size = 1000			
α_{11}	94.6	87.7	95.0	91.6	88.2	85.5
α_{12}	92.1	84.2	95.1	92.1	85.6	81.2
α_{13}	95.1	93.1	94.7	93.9	92.8	94.1
α_{14}	94.6	91.8	95.2	94.8	95.1	95.7
α_{21}	93.4	81.7	93.8	88.1	85.1	75.7
α_{22}	90.6	78.0	94.7	88.6	83.3	70.9
α_{23}	93.6	88.6	94.5	91.7	89.7	88.8
α_{24}	94.6	86.6	95.4	91.5	92.9	92.0

Table 5.7 Power of 2-sided test for $H_o : \alpha_{1i}$, $i = 1, 3$ and $H_o : \alpha_{2j}$, $j = 1, 2$

	Normal f_t, normal ϵ_t		Uniform f_t, uniform ϵ_t		χ^2 f_t, normal ϵ_t	
	PL	WLS	PL	WLS	PL	WLS
α_{11}	93.0	82.2	93.2	83.5	85.3	80.9
α_{13}	93.3	85.1	91.8	83.7	86.7	88.1
α_{21}	96.8	89.7	94.5	89.7	89.0	86.8
α_{22}	93.7	89.8	92.7	88.0	89.2	88.1

We also studied the performance of the PL procedure in small sample situations. For this, 1000 samples of size 150 were generated from the model (5.1) with mixed heteroscedastic error variance structure (5.3). The performance of the PL method was very good. As an example, we report the results for Case A. The coverage probabilities of the c.i.'s were 90.9%, 90.7%, and 89.3% for $\boldsymbol{\beta}_1$, and coverage probabilities for α_{1i} and α_{2i} ranged between 86.6% and 98% with an average of 93.9%. The power of the 2-sided test for $H_0 : \alpha_{1i} = 0, i = 1, 3$, and $H_0 : \alpha_{2j} = 0, j = 1, 2$, were high, ranging from 95.5% to 96.7% with an average of 96.3%.

The performance of the PL procedure was also assessed under situations where we have non-i.i.d. $\mathbf{f}_t$. For the model (5.1) with mixed heteroscedastic error variance structure (5.3), and standard normal ϵ_{it}^0, we considered fixed f_t and first order auto-regressive f_t. 1000 samples of size 300 were generated for each case. The PL method performed very well for both types of f_t. We summarize the results with the numbers for autoregressive f_t in parenthesis. The coverage probabilities were 94.2% (91.6%), 92.1% (93.0%), and 92.4% (92.1%) for $\boldsymbol{\beta}_1$, and ranged between 88.7% and 95.7% (between 89.1% and 97.4%) with an average of 94.1% (93.2%) for α_{1i} and α_{2i}. The power of the 2-sided test for $H_0 : \alpha_{1i} = 0$, $i = 1, 3$, and $H_0 : \alpha_{2j} = 0$, $j = 1, 2$, were high, ranging from 93.3% to 98.0% (from 92.6% to 96.0%) with an average of 95.2% (94.3%). Thus, with its good performance across all factor types, the PL method should be useful in practice.

5.2 Simulation Study II

A simulation study was conducted to evaluate the various approaches to factor score estimation presented in Chapter 3.

Throughout, we used

$$\mathbf{Z}_t = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ 1 \end{pmatrix} f_t + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}, \tag{5.5}$$

$$\epsilon_{it} = \sqrt{\alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2}\, \epsilon_{it}^0,$$

where ϵ_{it}^0's are standard normal variates, and the true values of the $\boldsymbol{\beta}$-parameters are given by (5.2). For the $\boldsymbol{\alpha}$-parameters, we consider two models; the heteroscedastic with values as given in (5.4) and the homoscedastic with $\boldsymbol{\alpha}_0$ in (5.4) and $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \mathbf{0}$. For sample size, we used n=1000. The number of simulated samples was 1000 for each model. The true factors f_t's were generated in the following manner. First, a fixed initial set of 1000 values of f_t's was generated from a Normal(1.5, 0.75) distribution. This set of 1000 true values of f_t's was used for all simulated samples in both models. Thus, the simulation treats f_t as fixed. This was appropriate for assessing performance of estimators of f_t over simulated samples. The ϵ_{it}^0's were generated independently for every sample.

For estimation, we considered two scenarios

a. All parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and all error variances Ψ_t are known,

b. $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ $\boldsymbol{\alpha}$, and Ψ_t are unknown.

In Scenario a, two sets of factor score estimates were computed; the homoscedastic factor estimates $\tilde{f}_{hom,t}$ in (3.21) and the heteroscedastic estimates $\hat{f}_{het,t}^{(2)}$ in (3.27), both using the true values of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, Ψ_0 and Ψ_t. In Scenario b, we obtained $\tilde{f}_{hom,t}$ using the homoscedastic estimates of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and Ψ_0, and two values of $\hat{f}_{het,t}^{(2)}$ with the PL and WLS estimates of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\alpha}$. For the variance estimation, $\tilde{V}_{HOM}$ in (3.22) and $\widehat{V}_{HOM,t}$ in (3.24) for $\tilde{f}_{hom,t}$, and $\widehat{Var}(\hat{f}_{het,t}^{(2)})$ in (3.28) for $\hat{f}_{het,t}^{(2)}$ are considered. For Scenario a, these variances are evaluated at the true values of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, Ψ_0 and Ψ_t. For Scenario b, the variances are evaluated at the corresponding estimates. Using these variance estimates, we considered three types of nominal 95% c.i. for f_t

$$\tilde{f}_{hom,t} \pm 1.96\sqrt{\tilde{V}_{HOM}}, \tag{5.6}$$

$$\tilde{f}_{hom,t} \pm 1.96\sqrt{\widehat{V}_{HOM,t}}, \tag{5.7}$$

$$\hat{f}^{(2)}_{het,t} \pm 1.96\sqrt{\widehat{Var}(\hat{f}^{(2)}_{het,t})}, \tag{5.8}$$

denoted by HOM, HOM-NEW and HET c.i.'s respectively.

We first report the results under the heteroscedastic true model. Since it is not plausible to present the estimation results for all 1000 f_t's, some crude criteria was set up to select a subset to be included in this discussion. Consider

$$k_{it} = \frac{\psi_{ii,t}}{\psi_{ii,0}}$$

where $\psi_{ii,t}$ is the true value of i^{th} diagonal element of Ψ_t and $\psi_{ii,0}$ is the true value of i^{th} diagonal element of Ψ_0. Note that k_{it} can be thought of as a crude measure of heteroscedasticity and ranges in value from about 0.3 to 2.5. Values of k_{it} farther from unity would indicate more severe error heteroscedasticity. Because of the true parameter values used in this study, k_{it}, $i = 1, 2, 3, 4$, have very strong positive correlation, well in excess of 0.9. Hence it is reasonable to use k_{1t} in lieu of the others. Based on the values of k_{1t}, 12 of the factor values were chosen and the estimation results for these 12 f_t's are presented.

Figure 5.8 gives the MSE of $\tilde{f}_{hom,t}$ and $\hat{f}^{(2)}_{het,t}$ for Scenario a (the 12 values are connected by lines). The MSE of $\hat{f}^{(2)}_{het,t}$ (HET) is consistently smaller than that of $\tilde{f}_{hom,t}$ (HOM), and the difference grows with the level of severity in error heteroscedasticity as measured by k_{1t}. The average MSE is 0.092 for $\hat{f}^{(2)}_{het,t}$ and 0.098 for $\tilde{f}_{hom,t}$, which is about 6% larger than the former.

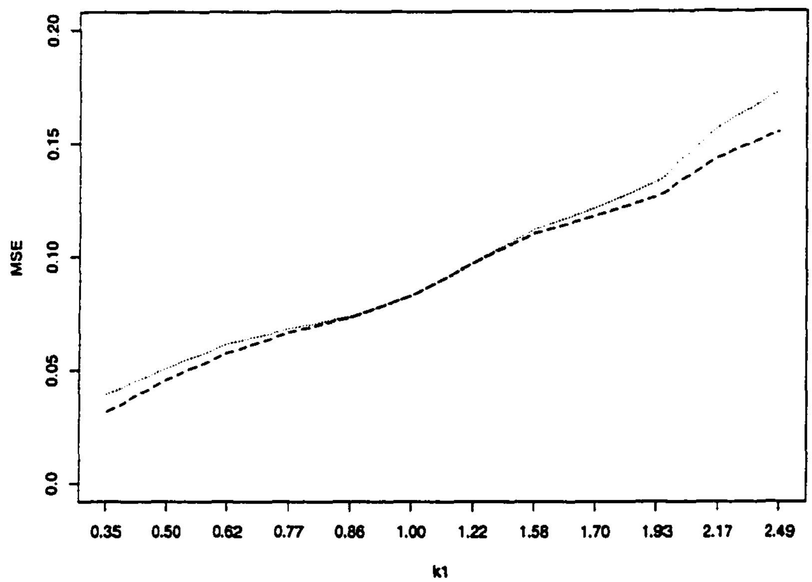

Figure 5.8 Scenario a - Comparison of MSE
(heteroscedastic data)
.... HOM - - - HET

Figure 5.9 plots the empirical coverage probabilities of the c.i.'s (5.6), (5.7), and (5.8). Note that for this Scenario a, the variance estimates in (5.6), (5.7), and (5.8) are true values. Clearly, HOM does not provide acceptable coverage probability, except when k_{1t} is close to 1. As the value of k_{1t} departs from 1, the coverage probability of HOM worsens. Using the individualized variance estimate $\widehat{V}_{HOM,t}$ for $\tilde{f}_{hom,t}$, the coverage probability improves dramatically (see HOM-NEW). On the other hand, HET which also uses the information on heteroscedasticity has excellent coverage probabilities as well.

In Scenario b, all the parameters and error variances are unknown and need to be estimated. We have the set of homoscedastic factor estimate $\tilde{f}_{hom,t}$, and two sets of heteroscedastic factor estimates $\hat{f}^{(2)}_{PL,t}$, $\hat{f}^{(2)}_{WLS,t}$ based on $\hat{f}_{het,t}$ and the PL and WLS model fit. Figure 5.10 shows the MSE of $\tilde{f}_{hom,t}$, $\hat{f}^{(2)}_{PL,t}$, and $\hat{f}^{(2)}_{WLS,t}$, denoted by HOM, PL, and WLS respectively. The average MSE for HOM, PL, and WLS are 0.109, 0.123, and

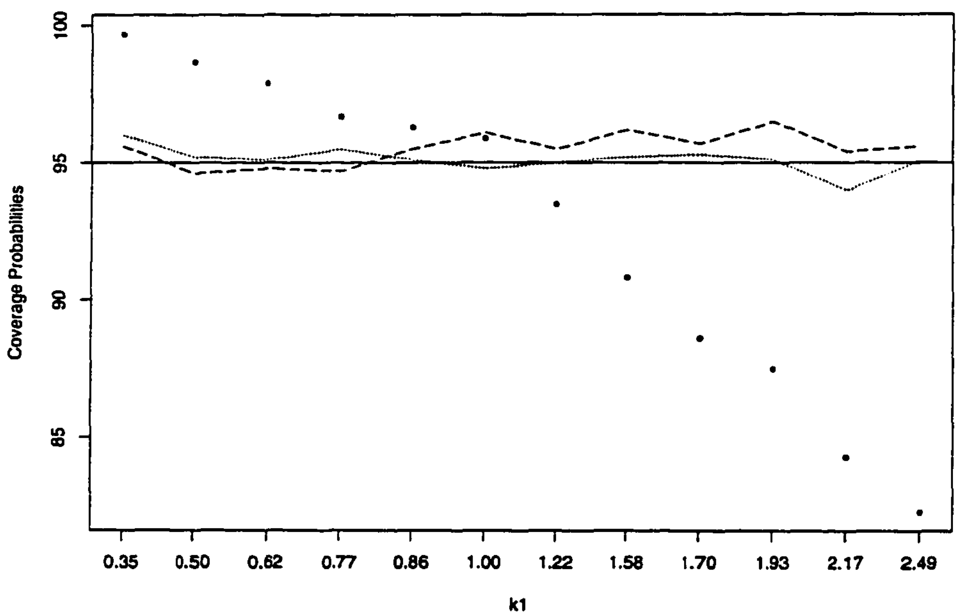

Figure 5.9 Scenario a - Comparison of coverage probabilities (heteroscedastic data)
• • • HOM .... HOM-NEW
– – – HET

0.274. $\hat{f}^{(2)}_{WLS,t}$ does not perform well in comparison to the other two estimators. As discussed in Chapter 3, the WLS approach involves more use of higher moments of $\mathbf{Z}_t$ and good estimates of these moments are difficult to secure. The volatility inherent in the estimation of these higher moments is manifested in the relatively poor estimates for the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and subsequently, the estimates for f_t.

For the variance estimation, $\tilde{V}_{HOM}$ in (5.6) is evaluated at the homoscedastic estimates of $\boldsymbol{\beta}_1$ and Ψ_0. For $\widehat{V}_{HOM,t}$ in (5.7), we considered two versions, $\widehat{V}_{HOM,PL,t}$ and $\widehat{V}_{HOM,WLS,t}$ using the homoscedastic estimate of $\boldsymbol{\beta}_1$ and Ψ_0, and the PL and WLS estimate of $\boldsymbol{\alpha}$. For $\widehat{Var}(\hat{f}^{(2)}_{het,t})$ in (5.8), we have two sets of variance estimates, corresponding to $\hat{f}^{(2)}_{PL,t}$ and $\hat{f}^{(2)}_{WLS,t}$. Thus we have five different c.i.'s, denoted by HOM (5.6), HOM-PL and HOM-WLS (5.7), as well as PL and WLS (5.8). Figure 5.11 shows the coverage

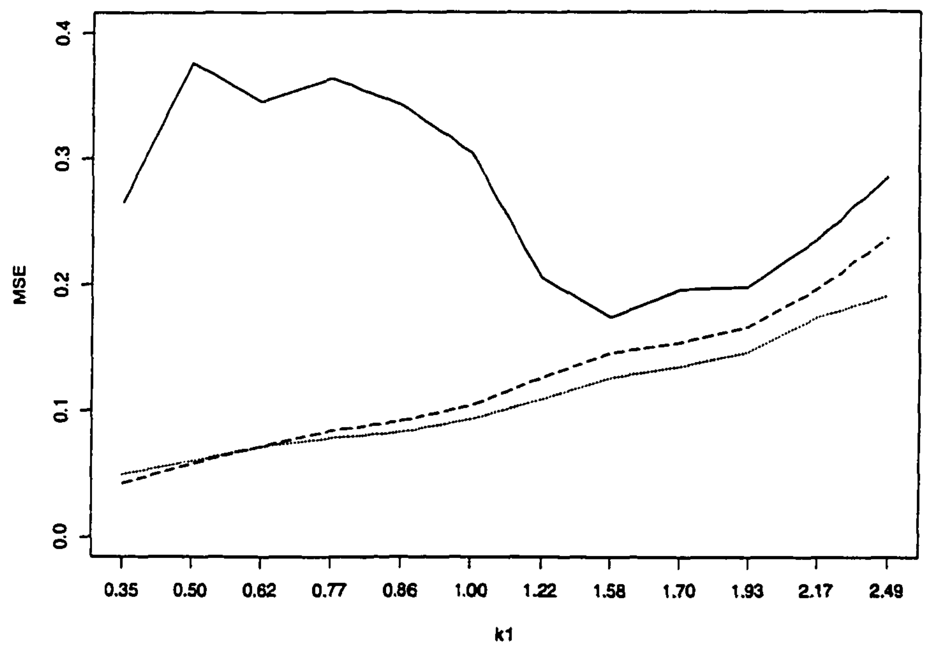

Figure 5.10 Scenario b - Comparison of MSE
(heteroscedastic data)
.... HOM — — — PL ——— WLS

probabilities of the five 95% c.i.'s. Again, HOM, yielded the worst coverage. The results improved greatly when the individualized variance estimator $\widehat{V}_{HOM,PL,t}$ or $\widehat{V}_{HOM,WLS,t}$ are used instead of $\tilde{V}_{HOM}$, with HOM-PL performing better than HOM-WLS. Also, PL and WLS were also better than HOM, but not as good as HOM-PL or HOM-WLS.

Now, we report the simulation results obtained under the homoscedastic model. Only Scenario b was considered. Also, based on the simulation study in Section 5.1, we did not include any methods involving the WLS estimation. For reporting, we selected a subset of 12 f_t out of the 1000, with true values ranging from 0.4 to 2.7.

Figure 5.12 shows the MSE of $\tilde{f}_{hom,t}$ and $\hat{f}^{(2)}_{PL,t}$, denoted as HOM and PL. The average MSE is 0.030 for $\tilde{f}_{hom,t}$ and 0.031 for $\hat{f}^{(2)}_{PL,t}$. As can be expected, the standard factor estimator which correctly assumes homoscedasticity has consistently lower MSE.

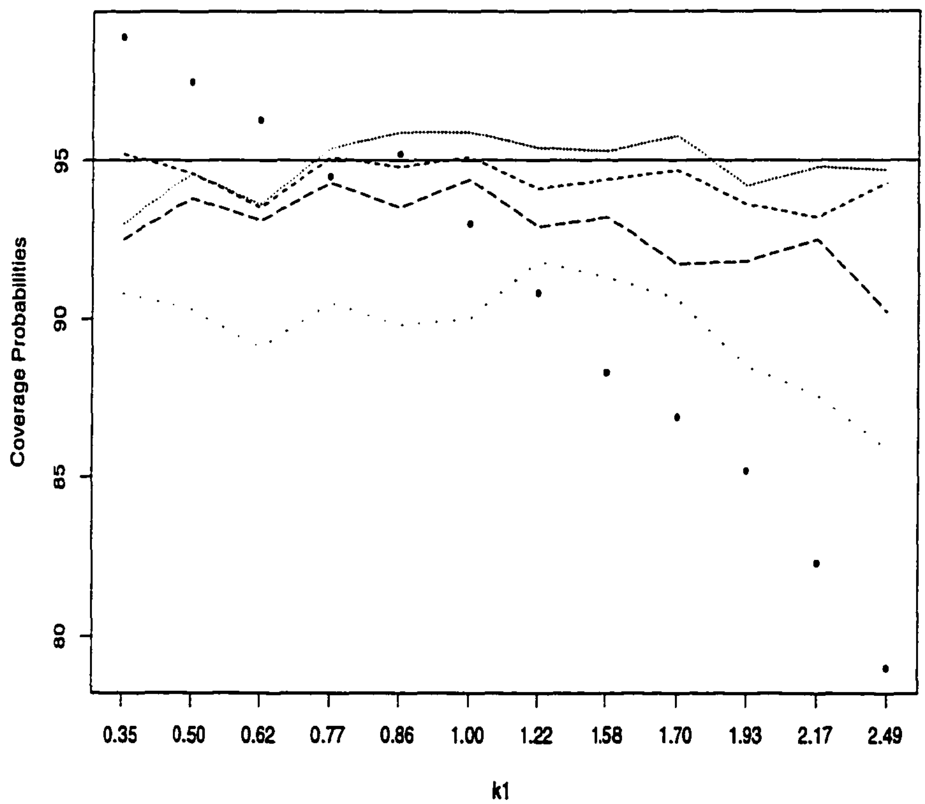

Figure 5.11 Scenario b - Comparison of coverage probabilities (heteroscedastic data)
• • • HOM .... HOM-PL
- - - - HOM-WLS – – – PL . . . . WLS

However, the difference between HOM and PL is very small. Hence, we do not lose much efficiency in using the heteroscedastic PL factor estimator, even when the true model is homoscedastic.

Figure 5.13 shows the coverage probabilities of the 95% c.i.'s for HOM, HOM-PL, and PL. The coverage probabilities are all similarly close to the 95% level. Given that all these c.i.'s have acceptable coverage level, we compared the inverval widths by comparing the averages (over 1000 simulated samples) of the estimated standard errors.

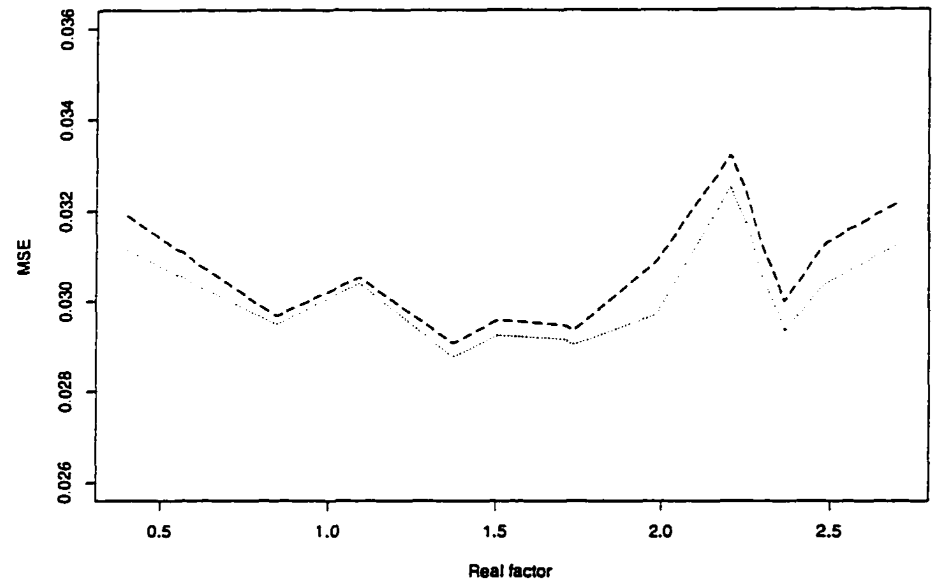

Figure 5.12 Comparison of MSE
...... HOM - - - - PL

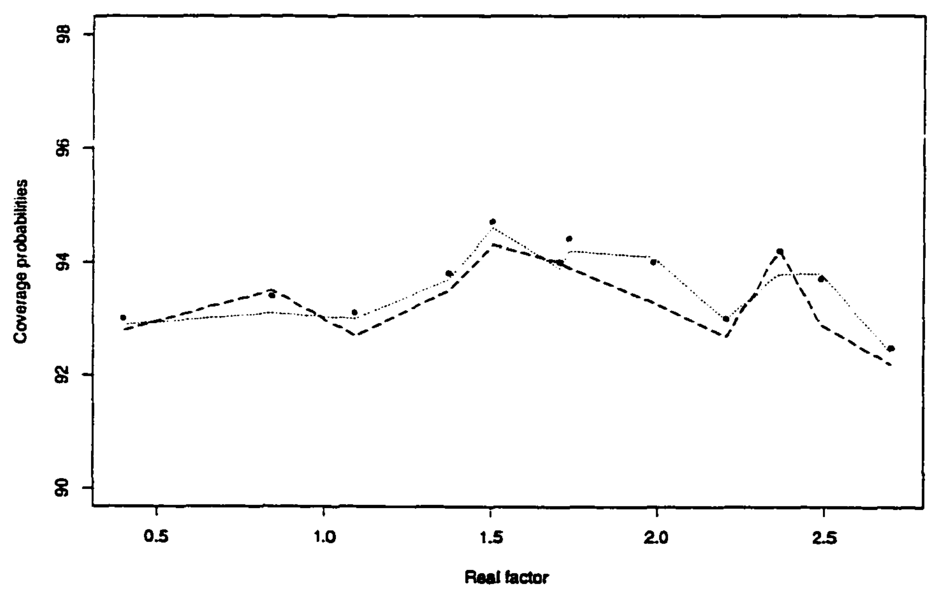

Figure 5.13 Comparison of coverage probabilities
● ● ● HOM .... HOM-PL - - - - PL

Figure 5.14 plots the three values SDHOM, SDHOM,t, and SDPL, which are the averages of $\sqrt{\tilde{V}_{HOM}}$, $\sqrt{\widehat{V}_{HOM,PL,t}}$, and $\sqrt{\widehat{Var}(\hat{f}^{(2)}_{PL,t})}$. As can be seen in Figure 5.14, the three standard errors are similar, but the PL c.i. is shorter than the two others, especially so for f_t values far away from the center. Thus, even when the true model is homoscedastic, the PL method can give informative c.i.'s without sacrificing the accuracy.

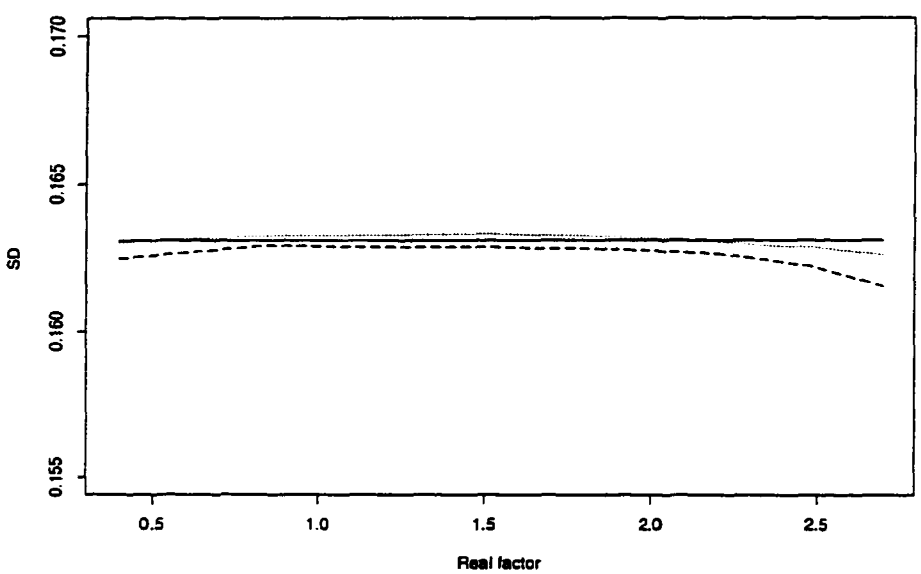

Figure 5.14 Comparison of standard errors
—— SDHOM .... SDHOM,t
- - - SDPL

5.3 Overall Recommendation

For estimation and inference for β_1, the standard homoscedastic procedures are robust against modest heteroscedasticity, and are useful for a wide range of distributions. But, the PL method can provide equally efficient and accurate inferences for β_1 regardless of the presence or absence of heteroscedasticity. In addition, the PL method can provide accurate and powerful tests for checking the heteroscedasticity. For the factor score estimation, the individual-specific variance estimation can be useful and valid for inference in practice. The use of the homoscedastic $\tilde{f}_{hom,t}$ with the individual-specific variance estimator $\hat{V}_{hom,PL,t}$ or the PL $\hat{f}^{(2)}_{PL,t}$ can provide accurate and efficient inference regardless of the degree of heteroscedasticity.

In general, we recommend a preliminary examination of the data using the diagnostic plots suggested in Chapter 2. The factor estimates and residual estimates used in these plots can be obtained easily from a homoscedastic fit to the data. If any systematic pattern is found in the plots, a heteroscedastic model with polynomial error variances should be fitted, and a formal test for heteroscedasticity can then be conducted for each error element. For fitting the heteroscedastic model and testing, we recommend the PL method.

APPENDIX. SOME RESULTS

In this appendix we present some mathematical results referred to in Chapter 3. To fit the heteroscedastic factor model using the distance minimizing methods presented in Chapter 3, we needed to find the expectation of the first two moments of the augmented variable $\mathbf{Z}_{a,t} = (\mathbf{Z}_t', \mathbf{U}_t')'$. The explicit expressions for these expectations depends on the specific heteroscedastic error structure adopted and the additional variables included in $\mathbf{U}_t$. In Section 3.2, we discussed an example model to illustrate the idea of augmentation. The model was

$$\begin{aligned} \mathbf{Z}_t &= \begin{pmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \\ Z_{4t} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ 1 \end{pmatrix} f_t + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{pmatrix}, \qquad \text{(A.1)} \\ \epsilon_{it} &= g_i(f_t; \boldsymbol{\alpha})\, \epsilon_{it}^0, \\ g_i(f_t; \boldsymbol{\alpha}) &= \sqrt{\alpha_{0i} + \alpha_{1i} f_t + \alpha_{2i} f_t^2}. \end{aligned}$$

Here we derive the expectation of the first two moments of $\mathbf{Z}_{a,t}$ under this example model. Some of these expressions were used to obtain the model fits in the simulation studies in Chapter 5. The choice of variables suggested to be included in $\mathbf{U}_t$ were $W_{ij,t} = (Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j)$ and $Y_{it} = (Z_{it} - \bar{Z}_i)^2$. As will be shown, when $W_{ij,t}$'s are included in $\mathbf{U}_t$, these expectations have expressions which are polynomials in $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\alpha}$, and the first four factor moments. When Y_{it}'s are also included in $\mathbf{U}_t$, these expectations, except for $Var(Y_{it})$ and $Cov(W_{ij,t}, Y_{it})$, have expressions which are polynomials in $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$,

α, and the first four factor moments. Explicit expressions for $Var(Y_{it})$ and $Cov(W_{ij,t}, Y_{it})$ in terms of moments cannot be found. Hence, $W_{ij,t}$ is preferred over Y_{it} for fitting this model.

Depending on the choice of $W_{ij,t}$ and Y_{it} to be included in $\mathbf{U}_t$, sample mean $\bar{\mathbf{Z}}_a$ may involve quantities such as $\overline{W}_{ij}$ and $\bar{Y}_i$ in addition to $\bar{Z}_i$, where

$$\bar{Z}_i = \frac{1}{n}\sum_{t=1}^{n} Z_{it},$$

$$\overline{W}_{ij} = \frac{1}{n}\sum_{t=1}^{n} W_{ij,t},$$

and

$$\bar{Y}_i = \frac{1}{n}\sum_{t=1}^{n} Y_{it},$$

and the sample covariance of $\mathbf{Z}_{a,t}$, $\mathbf{S}_a$ may involve variance terms such as $S_{w_{ij}w_{ij}}$ and $S_{y_iy_i}$ and covariance terms such as $S_{z_iw_{jk}}$, $S_{z_iy_j}$, $S_{w_{ij}w_{kl}}$, $S_{w_{ij}y_k}$ and $S_{y_iy_j}$ in addition to $S_{z_iz_i}$ and $S_{z_iz_j}$, where

$$S_{z_iz_i} = \frac{1}{n-1}\sum_{t=1}^{n}(Z_{it} - \bar{Z}_i)^2,$$

$$S_{z_iz_j} = \frac{1}{n-1}\sum_{t=1}^{n}(Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j),$$

$$S_{w_{ij}w_{ij}} = \frac{1}{n-1}\sum_{t=1}^{n}(W_{ij}, -\overline{W}_{ij})^2,$$

$$S_{z_iw_{jk}} = \frac{1}{n-1}\sum_{t=1}^{n}(Z_{it} - \bar{Z}_i)(W_{jk}, -\overline{W}_{jk}),$$

$$S_{w_{ij}w_{kl}} = \frac{1}{n-1}\sum_{t=1}^{n}(W_{ij}, -\overline{W}_{ij})(W_{kl}, -\overline{W}_{kl}),$$

$$S_{y_iy_i} = \frac{1}{n-1}\sum_{t=1}^{n}(Y_{it} - \bar{Y}_i)^2,$$

$$S_{z_iy_j} = \frac{1}{n-1}\sum_{t=1}^{n}(Z_{it} - \bar{Z}_i)(Y_{jt} - \bar{Y}_j),$$

$$S_{w_{ij}y_k} = \frac{1}{n-1}\sum_{t=1}^{n}(W_{ij}, -\overline{W}_{ij})(Y_{kt} - \bar{Y}_k),$$

and

$$S_{y_i y_j} = \frac{1}{n-1}\sum_{t=1}^{n}(Y_{it}-\bar{Y}_i)(Y_{jt}-\bar{Y}_j).$$

Under model (A.1), these quantities have expectations

$$\begin{aligned}
E(\bar{Z}_i) &= \beta_{0i}+\beta_{1i}\mu_f, \\
E(\bar{W}_{ij}) &= E(W_{ij,t}) \\
&= E\left\{(Z_{it}-\bar{Z}_i)(Z_{jt}-\bar{Z}_j)\right\} \\
&\approx E\left\{(Z_{it}-E(Z_{it}))(Z_{jt}-E(Z_{jt}))\right\} \\
&= E\left\{(\beta_{1i}(f_t-\mu_f)+\epsilon_{it})(\beta_{1j}(f_t-\mu_f)+\epsilon_{jt})\right\} \\
&= \beta_{1i}\beta_{1j}\phi^2,
\end{aligned}$$

$$\begin{aligned}
E(\bar{Y}_i) &= E(Y_{it}) \\
&= E\left\{(Z_{it}-\bar{Z}_i)^2\right\} \\
&\approx E\left\{(Z_{it}-E(Z_{it}))^2\right\} \\
&= E\left\{(\beta_{1i}(f_t-\mu_f)+\epsilon_{it})^2\right\} \\
&= \beta_{1i}^2\phi^2+E(g_i^2(f_t;\boldsymbol{\alpha})),
\end{aligned}$$

$$\begin{aligned}
E(S_{z_i z_i}) &= Var(Z_{it}) \\
&= Var(\beta_{1i}f_t+\epsilon_{it}) \\
&= \beta_{1i}^2\phi^2+E(g_i^2(f_t;\boldsymbol{\alpha})),
\end{aligned}$$

$$\begin{aligned}
E(S_{z_i z_j}) &= Cov(Z_{it},Z_{jt}) \\
&= Cov(\beta_{1i}f_t+\epsilon_{it},\beta_{1j}f_t+\epsilon_{jt}) \\
&= \beta_{1i}\beta_{1j}\phi^2,
\end{aligned}$$

$$
\begin{aligned}
E(S_{z_i w_{jk}}) &= Cov(Z_{it}, W_{jk}) \\
&= Cov(\beta_{1i} f_t + \epsilon_{it}, (\beta_{1j}(f_t - \mu_f) + \epsilon_{jt})(\beta_{1k}(f_t - \mu_f) + \epsilon_{kt})) \\
&= E\left[(\beta_{1i}(f_t - \mu_f) + \epsilon_{it})(\beta_{1j}(f_t - \mu_f) + \epsilon_{jt})(\beta_{1k}(f_t - \mu_f) + \epsilon_{kt})\right] \\
&= \begin{cases} \beta_{1i}\beta_{1j}\beta_{1k} E(f_t - \mu_f)^3, & \text{if } i \neq j \neq k \\ \beta_{1i}^2 \beta_{1k} E(f_t - \mu_f)^3 + \beta_{1k} E((f_t - \mu_f) g_i^2(f_t; \boldsymbol{\alpha})), & \text{if } i = j \neq k, \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
E(S_{w_{ij} w_{ij}}) &= Var(W_{ij,t}) \\
&= E\left\{(W_{ij,t} - E(W_{ij,t}))^2\right\} \\
&= E\left\{\left[(\beta_{1i}(f_t - \mu_f) + \epsilon_{it})(\beta_{1j}(f_t - \mu_f) + \epsilon_{jt}) - \beta_{1i}\beta_{1j}\phi^2\right]^2\right\} \\
&= \beta_{1i}^2 \beta_{1j}^2 \left[E(f_t - \mu_f)^4 - \phi^4\right] + \beta_{1i}^2 E\left[(f_t - \mu_f)^2 \epsilon_{jt}^2\right] \\
&\quad + \beta_{1j}^2 E\left[(f_t - \mu_f)^2 \epsilon_{it}^2\right] + E(\epsilon_{it}^2 \epsilon_{jt}^2) \\
&= \beta_{1i}^2 \beta_{1j}^2 \left[E(f_t - \mu_f)^4 - \phi^4\right] + \beta_{1i}^2 E\left[(f_t - \mu_f)^2 g_j^2(f_t; \boldsymbol{\alpha})\right] \\
&\quad + \beta_{1j}^2 E\left[(f_t - \mu_f)^2 g_i^2(f_t; \boldsymbol{\alpha})\right] + E(g_i^2(f_t; \boldsymbol{\alpha}) g_j^2(f_t; \boldsymbol{\alpha})),
\end{aligned}
$$

$$
\begin{aligned}
E(S_{w_{ij} w_{kl}}) &= Cov(W_{ij,t}, W_{kl,t}) \\
&= Cov\Big((\beta_{1i}(f_t - \mu_f) + \epsilon_{it})(\beta_{1j}(f_t - \mu_f) + \epsilon_{jt}), \\
&\qquad\qquad (\beta_{1k}(f_t - \mu_f) + \epsilon_{kt})(\beta_{1l}(f_t - \mu_f) + \epsilon_{lt})\Big) \\
&= E\Big\{\beta_{1i}\beta_{1j}\beta_{1k}\beta_{1l}\left[(f_t - \mu_f)^2 - \phi^2\right]^2 \\
&\qquad + \beta_{1i}\beta_{1k}(f_t - \mu_f)^2 \epsilon_{jt}\epsilon_{lt} + \beta_{1i}\beta_{1l}(f_t - \mu_f)^2 \epsilon_{jt}\epsilon_{kt} \\
&\qquad + \beta_{1j}\beta_{1k}(f_t - \mu_f)^2 \epsilon_{it}\epsilon_{lt} + \beta_{1j}\beta_{1l}(f_t - \mu_f)^2 \epsilon_{it}\epsilon_{kt}\Big\} \\
&= \begin{cases} \beta_{1i}\beta_{1j}\beta_{1k}\beta_{1l}(E(f_t - \mu_f)^4 - \phi^4), & \text{if } \{i,j\} \cap \{k,l\} = \emptyset \\ \beta_{1i}^2 \beta_{1j}\beta_{1k}\beta_{1l}(E(f_t - \mu_f)^4 - \phi^4) & \\ \quad + \beta_{1j}\beta_{1k} E[(f_t - \mu_f)^2 g_i^2(f_t; \boldsymbol{\alpha})], & \text{if } i = l, j \neq k, \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
E(S_{z_i y_j}) &= Cov(Z_{it}, Y_{jt}) \\
&= Cov(\beta_{1i}(f_t - \mu_f) + \epsilon_{it}, (\beta_{1j}(f_t - \mu_f) + \epsilon_{jt})^2) \\
&= E[(\beta_{1i}(f_t - \mu_f) + \epsilon_{it})(\beta_{1j}^2(f_t - \mu_f)^2 + \epsilon_{jt}^2 + 2\beta_{1j}(f_t - \mu_f)\epsilon_{jt})] \\
&= \beta_{1i}^3 E(f_t - \mu_f)^3 + \beta_{1i} E((f_t - \mu_f) g_j^2(f_t; \boldsymbol{\alpha})),
\end{aligned}
$$

$$
\begin{aligned}
E(S_{y_i y_i}) &= Var(Y_{it}) \\
&= E(\beta_{1i}(f_t - \mu_f) + \epsilon_{it})^4 - [E(\beta_{1i}(f_t - \mu_f) + \epsilon_{it})^2]^2 \\
&= E(\beta_{1i}^4(f_t - \mu_f)^4) + 6E(\beta_{1i}^2(f_t - \mu_f)^2\epsilon_{it}^2) + 4E(\beta_{1i}(f_t - \mu_f)\epsilon_{it}^3) + E(\epsilon_{it}^4) \\
&\quad - \left[E(\beta_{1i}^2(f_t - \mu_f)^2) + E(\epsilon_{it}^2)\right]^2 \\
&= \beta_{1i}^4 E((f_t - \mu_f)^4) + 6\beta_{1i}^2 E((f_t - \mu_f)^2 g_i(f_t; \boldsymbol{\alpha})) \\
&\quad + 4\beta_{1i} E((f_t - \mu_f) g_i^3(f_t; \boldsymbol{\alpha})) E(\epsilon_{it}^{03}) + E(\epsilon_{it}^{04}) \\
&\quad - \beta_{1i}^4\phi^4 - E(g_i^2(f_t; \boldsymbol{\alpha})^2) - 2\beta_{1i}^2\phi^2 E(g_i^2(f_t; \boldsymbol{\alpha})) - [E(g_i^2(f_t; \boldsymbol{\alpha}))]^2,
\end{aligned}
$$

$$
\begin{aligned}
E(S_{y_i y_j}) &= Cov(Y_{it}, Y_{jt}) \\
&= Cov((\beta_{1i}(f_t - \mu_f) + \epsilon_{it})^2, (\beta_{1j}(f_t - \mu_f) + \epsilon_{jt})^2) \\
&= \beta_{1i}^2\beta_{1j}^2 Var((f_t - \mu_f)^2) + Cov(\epsilon_{it}^2, \epsilon_{jt}^2) \\
&= \beta_{1i}^2\beta_{1j}^2((f_t - \mu_f)^4 - \phi^4) + E(g_i^2(f_t; \boldsymbol{\alpha}) g_j^2(f_t; \boldsymbol{\alpha})) \\
&\quad - E(g_i^2(f_t; \boldsymbol{\alpha})) E(g_j^2(f_t; \boldsymbol{\alpha})),
\end{aligned}
$$

$$
\begin{aligned}
E(S_{w_{ij} y_k}) &= Cov(W_{ij,t}, Y_{kt}) \\
&= Cov((\beta_{1i}(f_t - \mu_f) + \epsilon_{it})(\beta_{1j}(f_t - \mu_f) + \epsilon_{jt}), (\beta_{1k}(f_t - \mu_f) + \epsilon_{kt})^2) \\
&= \begin{cases}
\beta_{1i}^3\beta_{1j}(E(f_t - \mu_f)^4 - \phi^4) & \\
\quad + \beta_{1i}\beta_{1j} E\left\{(3(f_t - \mu_f)^2 - \phi^2) g_i^2(f_t; \boldsymbol{\alpha})\right\} & \\
\quad + \beta_{1j} E((f_t - \mu_f) g_i^3(f_t; \boldsymbol{\alpha}), & \text{if} \quad i = k \\
\beta_{1i}^2\beta_{1j}\beta_{1k}(E(f_t - \mu_f)^4 - \phi^4) & \\
\quad + \beta_{1j}\beta_{1k} E\left\{((f_t - \mu_f)^2 - \phi^2) g_i^2(f_t; \boldsymbol{\alpha})\right\}, & \text{if} \quad i \neq k.
\end{cases}
\end{aligned}
$$

Given explicit expressions for $E(g_i^2(f_t;\boldsymbol{\alpha}))$, $E(g_i^2(f_t;\boldsymbol{\alpha})g_j^2(f_t;\boldsymbol{\alpha}))$, $E((f_t-\mu_f)g_i^2(f_t;\boldsymbol{\alpha}))$, $E((f_t-\mu_f)^2g_i^2(f_t;\boldsymbol{\alpha}))$, and $E((f_t-\mu_f)g_i^3(f_t;\boldsymbol{\alpha}))$, a complete evaluation of these expectations can be made. Except for $E((f_t-\mu_f)g_i^3(f_t;\boldsymbol{\alpha}))$, valuation of all these quantities is straightforward under model (A.1).

BIBLIOGRAPHY

Amemiya, Y. and Fuller, W. A. (1984). Estimation for the multivariate errors-in0variables model with estimated error covariance matrix. *The Annals of Statistics* **12** 497-509.

Amemiya, Y. and Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics* **18** 1453-1463.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

Anderson, T. W. and Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics* **16** 759-771.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* Wiley, New York.

Browne. M. W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology* **41** 193-208.

Demos, A. and Sentana, E. (1998). An EM algorithm for conditionally heteroscedastic factor analysis. *Journal of Business and Economic Statistics* **16** 357-361.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Hasabelnaby, N. A. (1987). The Use of a weighting function in measurement error regression. Ph.D. dissertation, Iowa State University, Ames.

Jaccard, J. and Wan, C.K. (1996). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin* **117** 348-57.

Jöreskog, K. G. and Yang, F. (1996). Non-linear structural equation models: The Kenny-Judd model with interaction effects. In G.A. Marcoulides and R.E. Schumacker (Eds.). *Advanced Structural Equation Modelling: Issues and Techniques.* Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Jöreskog, K. G. and Yang, F. (1997). Estimation of interaction models using the augmented moment matrix: Comparison of asymptotic standard errors. In W. Bandilla and F. Faulbaum (Eds.) *SoftStat '97 Advances in Statistical Software 6.* Lucias & Lucias, Stuttgart.

Kenny, D. and Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin* **96** 201-210.

Lawley, D. N. and Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method.* Butterworths, London.

Meijer, E. and Mooijaart, A. (1996). Factor analysis with heteroscedastic errors. *British Journal of Mathematical and Statistical Psychology* **49** 189-202.

Ping, R. A. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin* **119** 166-175.

Sanger, T. M. and Fuller, W. A. (1991). Estimation of the measurement error model with unequal error variances. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 146-151.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology.* Blackwell Publishers **22** 249-278.

Wall, M. and Amemiya, Y. (1998). Generalized appended product indicator procedure for nonlinear structural equation analysis. Research Report 98-015, Division of Biostatistics, University of Minnesota.